A STUDY ON MIXTURE OF SOME UNIVARIATE DISCRETE PROBABILITY DISTRIBUTIONS

A
Thesis
Submitted to
Tezpur University

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in
Mathematical Sciences (Statistics)

By

ARUNDHATI DEKA NATH
(Regn. No 039 of 1999)

Department of Mathematical Sciences
Under The School of Science and Technology
TEZPUR UNIVERSITY
Napaam, Tezpur

2002

TEZPUR UNIVERSITY

(A Central University established by an Act of Parliament)

NAPAAM, TEZPUR - 7840028

DISTRICT : SONITPUR :: ASSAM :: INDIA

Ph : 03712 67004, 67005, 67006, 67007, 67008, 67009

Fax : 03712 - 67006, 03712 - 67005

e-mail : adm@agnigarh.tezu.ernet.in

Dt. 15/05/02

TO WHOM IT MAY CONCERN

It is certified that the research work in the thesis entitled " A study on Mixture of Some Univariate Discrete Probability Distributions " submitted by Ms Arundhati Deka Nath for the degree of Doctor of Philosophy (Statistics) was carried out under my supervision. The results presented here are of her own work and the thesis or part of it has not been submitted to any other university or institute for any research degree.

She has fulfilled all the requirements under rules and regulations for the award of the degree of Doctor of Philosophy of Tezpur University, Napaam, Tezpur.

Deptt of Mathematical Science, Tezpur University, TEZPUR ASSAM

M Borah 30/5/2002

(Dr.M.borah)
Reader
Department of Mathematical sciences

Acknowledgement

With immense pleasure, the authoress takes the privilege to express her deepest sense of gratitude and indebt ness to Dr. Munindra Borah, Reader, Department of Mathematical Sciences, Tezpur University, Tezpur for his keen interest, enthusiastic encouragement, illuminating guidance and highly esteemed suggestions during the whole course of the investigation and preparation of this manuscript.

The authoress is extremely grateful to Dr. A.K. Borkakoty, Professor & Head of the Department of Mathematical Sciences, Tezpur University, Tezpur for his valuable suggestion and other faculty members for their encouragement. The authoress is also deeply obliged and thankful to the Department of Computer Sciences for providing computer facility for carrying out the initial investigation.

The authoress expresses her most sincere gratitude to her respected mentor Dr. Bhubaneshwar Saharia, Academic Register, Tezpur University, Tezpur for his scholarly suggestion and continuous encouragement.

She acknowledges her gratitude to her teacher Dr. H.K.Baruah, Professor Department of Statistics, Gauhati University for giving the valuable knowledge of computer programming during her master degree course which helps her immensely during this investigation.

She is also extremely grateful to Dr. Abani Kumar Sarma, Principal, Darrang College, Tezpur for his positive support and his kind help through out the investigation

She extends her warm thanks and gratefulness to Dr.B.N.Pandey, Department of Statistics, Banaras Hindu University for who's strong recommendation she was honored to receive the invitation from the Organizing Committee, Sixth International Conference of the Forum for Interdisciplinary Mathematics, University of South Alabama, U.S.A.

The authoress is extremely grateful to Dr. J.Roy and Dr. Bikas Sinha of ISI, Calcutta and K.G.Janardan, Eastern Michigan University, U.S.A. for their encouragement and advice at some critical spheres of the investigation. She is also extremely thankful to Prof. C.R. Rao for his valuable comment on one of her research paper presented in the 4th International Triennial Calcutta Symposium on Probability & Statistics, Dec. 26-28, 2000.

She expresses her gratefulness to the publishers and the editorial boards, those who have timely published her papers during the course of investigation.

No words could be described the gratitude and sense of obligation she feels for her dear family, for showing their blessings, love and support and for being the impetus behind all her endeavors. She asserts her deep mindful indebt ness and sense of obligation to Er. Utpal Nath for being a fort of strength and support in every steps of this investigation.

She is extremely thankful to all her colleagues, friends and juniors for their constant inspiration.

Above all she remembers almighty God for His benevolence and for synchronizing everything in perfect harmony which led to the successful accomplishment of this manuscript.

Dated, Tezpur
the 9th May, 2002

The Authoress

Preface

Investigations reported in this thesis entitled – 'A study on Mixture of some Univariate Discrete Probability Distributions' is undertaken by the author under the guidance of Dr. Munindra Bora, Reader, Department of Mathematical Sciences, Tezpur University, Tezpur.

This thesis contains mainly some finite and countable mixtures of univariate discrete probability distributions. The mixture of discrete probability distribution has become an extremely useful branch of statistics having important applications in a wide variety of disciplines such as biological and medical sciences, social sciences, physical sciences, operation research, quality control, engineering and so on. The moment one gets in to a stochastic problem where nothing more than counting is involved, one is dealing with discrete distribution. While pursuing this study I have been greatly influenced by learned works of several authors like Prof. P.C. Consul, G.C. Khatri, S.K. Katti, C.D. Kemp, A.W. Kemp, K.G. Janardan, G.P. Patil, L.R. Shenton, N.L. Johnson, S. Kotz, G.C. Jain, R.C. Gupta and I.G. Plunkett. I am indebted to all of them.

The works in this thesis are divided into seven chapters. The first chapter is an introductory one. It is devoted to the various techniques of mixture distributions that have been studied in this thesis. It also contains review of previous works and the synopsis of the thesis. In the rest of the chapters, we studied certain mixture distributions such as – inflated distributions, generalized distributions and Lagrangian type distributions. The parameters of each of the distribution are estimated and empirical fits are given to test the relative efficiency of different method of estimation. The whole study involves a lot of computer programming. The recurrence relations for probabilities and moments are aimed to derive in such a way that they are easy to handle on computers.

During the course of the investigation, I got the opportunities to attend certain international conferences where I got the chances to meet many renowned statisticians, specially- Prof.C.R.Rao, Prof.J.Roy, Prof.S.B.Sinha, Prof.G.P.Bhattacharya, Prof.N.R.Mohan etc. Their valuable advices encourage me to study further in this field.

In this thesis, due to limitation of spaces, only the important results are given though a large number of data sets were investigated during the course of investigation.

A few words about the notation used in this thesis are given below.

An equation is marked as c.s.n., where c stands for the 'chapter', s for the 'section' in which it occurs and n is the 'serial number'.

Similarly, the graphs and tables are marked as c.n., where c stands for 'chapter' and n is the 'serial number'.

Lastly, the published papers of author are given in the appendix for ready references.

INDEX

CHAPTER 1

- INTRODUCTION
- REVIEW ON PREVIOUS WORKS
- SYNOPSIS

Chapter 1

1.1 Introduction

The theory of discrete probability distribution is an extremely useful branch of statistics having important applications in a wide variety of disciplines. The origin of this theory began with the work of James Bernoulli (1713) and Poisson (1837). In recent years the mixture distributions have received continuous attention since elementary distribution such as Poisson, binomial, logarithmic which can be formulated on the basis of simple models, have been found to be inadequate to describe the situations which occurs in number of phenomenon. Hence univariate mixture distributions, which combine two or more of the elementary distributions through the process of compounding or generalizing, have become an extremely useful branch of statistics. These distributions have important applications in biological sciences, medical sciences, social sciences, physical sciences, operation research, engineering and so on. A detailed account of these discrete mixture distributions and their properties can be found in the works of Neyman (1939), Gurland (1957,1958,1965), Haight (1961), Patil (1961,1962a,1962b,1964), Khatri (1961,1962), Katti (1966), Katti and

Gurland (1961,1962a,1962b) and in the books of Johnson and Kotz (1969), Johnson et al (1992) and Consul (1989).

1.2 Mixture Distribution

A mixture distribution is a superimposition of distributions with different functional forms or different parameters, in specified proportions. Sometimes, however, mixing is just a mechanism for constructing a new distribution for which empirical justification is sought later on. If $F_j(x_1, x_2, \ldots x_n), (j = 0,1,\ldots,m)$ represents different cumulative distribution functions (cdf) and $a_j \geq 0$ and $\sum_{j=0}^{m} a_j = 1$ then

$$F(x_1, x_2, \ldots x_n) = \sum_{j=0}^{m} a_j F_j(x_1, \ldots x_n)$$

is a proper cumulative distribution function. This mixture distribution $\{F_j\}$ is finite or infinite according to m is finite or infinite. Thus two important categories of mixture distributions are finite mixture and countable or continuous mixture of discrete distributions.

(a) Finite mixture

The concept of finite mixture of distribution was introduced into literature by Pearson (1915). A mixture distribution is said to be a finite mixture of distribution if

$$F(x) = p_1 F_1(x) + \ldots\ldots p_k F_k(x)$$

where $p_1, p_2, ..., p_k$ are the weights of the compound distribution with cdfs $F_1(x), F_2(x), ..., F_k(x)$ and $p_j > 0$, $\sum_{j=1}^{k} p_j = 1$.

In finite mixture of distribution the problem of central interest arises when data are not available for each conditional distribution separately but available only for the overall mixture distribution. Often such situation arises because it is impossible to observe some underlying variables which split the observations into groups – only the combined distribution can be studied. In these circumstances interest often focuses on estimating the mixing proportions and on estimating the parameters in the conditional distributions. Zero modified or inflated distribution is an example of finite mixture of distribution.

Inflated Distribution

A random variable X is said to have the inflated distribution if its probability mass function (pmf) is defined by

$$P(X = x) = \begin{cases} \beta + \alpha p_0, & x = 0 \\ \alpha p_x, & x = 1,2,3,... \end{cases}$$

where p_j, (j=0,1,2,...) is the pmf of the original distribution without inflation and $\alpha + \beta = 1$. It is also possible to take β less than zero, provided $\beta + \alpha p_0 \geq 0$.

The probability generating function (pgf) of inflated distribution is $H(t) = \beta + \alpha G(t)$, where $G(t)$ is the pgf of original distribution without inflation.

The studies of discrete inflated distributions were initiated by Singh (1963) to consider the probabilistic description of such experiment where there is some 'inflation' of the probability at the point zero for Poisson distribution. Singh (1966) studied the inflated binomial distribution. The generalized inflated Poisson distribution was investigated by Pandey (1965) in the sense that the inflation of the distribution occurs at an arbitrary point 'l' (being the value of the random variable X).

Thus the random variable X is said to have generalized inflated distribution if its pmf is expressed as

$$P(X=x)=\begin{cases}\beta+\alpha p_x, & x=l\\ \alpha p_x, & x=0,1,\ldots,l-1,l+1,\ldots\end{cases}$$

(b) Countable and Continuous Mixture of Discrete Distribution

A mixture distribution also arises when the cumulative distribution function of a random variable depends on the parameters $\theta_1,\theta_2,\ldots,\theta_m$ and some (or all) of those parameters may vary. A mixture distribution of this type is represented by

$$F_A \underset{\Theta}{\wedge} F_B$$

where F_A is the original distribution and F_B is the mixing distribution. When Θ has a discrete distribution with probability p_i, (i=0,1,2,...) we call the outcome a countable mixture of discrete distribution. The pmf of the mixture is

$$P(X=x)=\sum_{i\geq 0} p_i P_i(x)$$

where $P_j(x) = F_j(x) - F_j(x-1)$

A continuous mixture of discrete distribution arises when a parameter corresponding to some features of a model for a discrete distribution can be regarded as a random variable taking continuous values. Greenwood and Yule (1920), Lundberg (1940) first studied the theory of countable and continuous mixture of discrete distribution.

In case of mixture distribution there are three important theorems derived by Gurland (1957), Levy (1957) and Maceda (1948).

Gurland's generalization Theorem [Gurland (1957)]

According to this theorem, a distribution with pgf of the form $G_1(G_2(z))$ will be called a F_1 distribution generalized by the generalizing F_2 distribution provided that $G_2(z/k\phi) = [G_2(z/\phi)]^k$

Symbolically, it will be represented by $F_1 \vee F_2$.

Levy's theorem [Feller (1957)]

If and only if discrete probability distribution on the non negative integers is infinitely divisible, then its pgf can be written as

$$G(z) = e^{\lambda(g(z)-1)}$$

where $\lambda > 0$ and g(z) is an another pgf.

Maceda's theorem [Maceda (1948)]

According to this theorem, mixing Poisson distributions using an infinitely divisible distribution yields a Poisson-stopped-sum distribution.

1.3 Review on Previous Works of Mixture of Discrete Distributions

An increasing amount of efforts have been made in the last few years in the area of discrete mixture distributions. Sometimes it is found that a simple distribution such as binomial, Poisson, negative binomial, logarithmic etc. fails to describe a set of data which leads to the belief that the model underlying the distribution has some of the characteristics of the generalized or mixture model. Thus further research was made to examine if any simpler mixture distribution will describe the data to a better degree of satisfaction. In this process a large number of discrete distributions were derived which are classified as generalized, modified and contagious distributions. A detailed accounts of these discrete mixture distributions and their properties can be found in the books of Johnson and Kotz (1969), Everitt and Hand (1981), Consul (1989) and Johnson et al (1992).

According to Smith (1985) finite mixture of distribution can be used in medicine, where the categories are disease states, in economics, where the categories are discontinuous forms of behaviours, in fisheries research, where the components are of different ages and in sedimentology where the categories are of mineral types. Again according to Titterington (1990) finite mixture distribution is used in speech recognition and in image analysis. Everitt and Hand (1981) studied finite mixture of distributions in their book and estimate the parameters by the method of moments and maximum likelihood

Inflated distribution studied by Singh (1963) and Pandey (1965) is an another example of finite mixture of distribution. Singh (1966) also investigated generalized inflated binomial distribution. He investigated that this distribution will be applicable in those cases where the single binomial describes the situation well except for the '*l*'th cell, which is inflated, i.e. there are some more numbers of observations with '*l*' that can be expected on the basis of single binomial. Grzegorska (1973) studied the inflated generalized power series distribution and obtained the recurrence relation for moments of this distribution. Patel (1975) investigated inflated at zero power series distribution. Sobich and Szynal (1974) and Lingappaiah (1977) obtained some properties of inflated distribution. Gerstenkorn (1979) established the recurrence relation for the moments about an arbitrary point of class of discrete inflated distributions. The same author also investigated the moment recurrence relations for the generalized inflated negative binomial, Poisson and geometric distribution. Kemp (1986) and Kemp and Kemp (1988) used maximum likelihood method to estimate the parameters of inflated Poisson and binomial distribution. A zero modified geometric distribution was studied by Holgate (1964) as a model for the length of residence of animals in a specified habitat. Williams (1947) introduced logarithmic with zero distribution. Chatfield (1969) used the probability generating function of this distribution as a model for stationary purchasing behavior. Khatri (1961) and Patil (1964) obtained it by mixing binomials.

Cohen (1963) considered a mixture distributions formed from a Poisson component and a binomial component. Dawid and Skene (1979) considered a mixture of multinomial distributions arising in a model of observed rating.

The works on countable and continuous mixture of discrete distribution were developed by the "accident proneness" theory of Greenwood and Yule (1920). In their model, an individual was assumed to have accidents at random, with an intensity θ, where θ is assumed to have a gamma distribution over the population of individuals. The number of accidents per individual is therefore a Poisson distribution with the value of its parameter θ conditional on a generalization of a gamma variable, which leads to have a negative binomial distribution.

Different mixtures of Poisson distributions where the mixing distributions are countable or continuous are discussed in details in the book by Johnson et al (1992).

Poisson mixture of Poisson distribution i.e. Neyman Type A distribution has often been used to describe plant distributions, especially when reproduction of the species produces clusters. Evans (1953) found that Neyman Type A gave good results for plant distribution. Martin and Katti (1965) fitted 35 data sets with a number of standard distributions and they found that those distributions have wide applicability. Cresswell and Froggatt (1963) derived the Neyman Type A distribution in context of bus driver accidents.

The Poisson Pascal distribution which is a Poisson mixture of negative binomial distribution was introduced in the context of the special distribution of plant by Skellam (1952). Katti and Gurland (1961) studied its properties, method of its estimation and derived it from an entomological model.

The Hermite distribution which is a Poisson mixture of Bernoulli distribution was studied by Kemp and Kemp (1965). Plunkett and Jain (1975) derived a new distribution known as Gegenbauer distribution by mixing the Hermite distribution with gamma distribution. Borah (1984) studied the probability and moment properties of Gegenbauer distribution and Medhi and Borah (1984) investigated the four parameter generalized Gegenbauer distribution and had used estimation via moment and ratio of first two frequencies and $\bar{x}$ and s^2.

The Polay-Aeppli distribution described by Polay (1930) arises in a model where the objectives occur in clusters and the number of clusters having a Poisson distribution, while the number of objects per cluster has the geometric distribution. Douglas (1965,1980) obtained an approximate formula for the probability of this distribution. This distribution is also the limiting form of Beall and Rescia's generalization of the Neyman Type A, B and C distribution.

Regarding mixtures of binomial distributions i.e. beta binomial distribution studied by Ishii and Hayakawa (1960), Poisson binomial distribution discussed by Skellam (1952) were discussed in details in the book by Johnson et al (1992). Recent further works on mixtures of binomial distributions are studied by Bowman et al (1992).

Lagrangian expression for the derivation of the probabilities of certain discrete distributions has been used for many years. Consul and Shenton (1972, 1973,1975) and their co-workers have studied systematically the technique for deriving the distributions and their properties. Lagrangian binomial distribution was obtained by Mohanty (1966). Jain and Consul (1971) derived an analogous Lagrangian negative binomial distribution. The Lagrangian Poisson distribution was obtained by Consul and Jain (1973) as a limiting form of the Lagrangian negative binomial distribution. A detailed study was made on the properties of Lagrangian Poisson distribution by Consul in his book (1989). Lagrangian Katz family of distributions was studied by Consul and Famoye (1996).

A very broad class of distribution, i.e. power series distributions which includes many of the common distributions was studied by Khatri (1959) and Patil (1961,1962). Gupta (1974) studied the modified power series distributions. Tripathi et al (1986) studied the incomplete moments of modified power series distribution. Grzegorska (1973) studied inflated generalized power series distribution and Patel (1975) investigated inflated at zero power series distribution. Patel (1975) obtained the maximum likelihood estimate of the parameters of inflated power series distribution.

1.4 Synopsis of the Thesis

The thesis entitled "A Study on Mixture of some Univariate Discrete Probability Distributions" comprises of seven chapters in all. The first chapter is an introductory one. It gives an account of the relevant works in the theory of univariate discrete probability distributions. The earlier works on different types of finite, countable and continuous mixture of some discrete distributions are discussed.

In the second chapter, inflated distribution with inflation of probability at zero and at an arbitrary point, say '*l*', are investigated. Inflated binomial, Poisson, negative binomial and geometric distributions are further investigated to study the recurrence relation for probabilities and moments of inflated at zero and generalized inflated distributions. The fitting of the inflated distributions are also considered. It has been seen that if there is an excess frequency of observed event at point zero as well as a respective decrease of its value at the remaining points, the inflated distribution provides much closer fit to the data than the classical one. The inflated power series distribution is also investigated. The recurrence relation for factorial moment and central moment are studied.

In chapter 3, two mixture distributions of Poisson Lindley distribution, namely, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distribution are obtained by using Gurland generalization (1957) theorem. Here, an attempt has been made to derive the recurrence relation for probabilities and moments without derivatives so that it will be easier to handle on computer. The parameters of these distributions are estimated by the method of moments and the ratio of first two frequencies with mean. A few reported data sets have been considered for empirical fitting of Poisson-Lindley, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distribution with remarkable results.

In chapter 4, inflated Poisson-Lindley distribution has been studied with some inflation at zero. Some properties of inflated Poisson-Lindley distribution are also discussed. The recurrence relation for probabilities as well as for moments and factorial moments are derived. The skewness and kurtosis of the distribution are studied. The parameters of the distribution have been estimated by the method of maximum likelihood, method of moments and the ratio of first two frequencies with mean. Different applications of inflated Poisson-Lindley distribution are discussed. The fits are compared with the generalized Poisson distribution with varied amount of success.

In chapter 5, an extension of Poisson-Poisson-Lindley distribution, i.e. short Poisson-Poisson-Lindley distribution is investigated. The model for derivation of the distribution has been discussed. The probability recurrence relation and moment recurrence relation are also studied. The application of this distribution has been considered. A few sets of accident data have been considered for fitting of the short Poisson-Poisson-Lindley distribution. The fits of short Poisson-Poisson-Lindley distribution are also compared with the fits given by the 'short' distribution and some other distributions as obtained by different authors and found much closer fit in all the cases for short Poisson-Poisson-Lindley distribution.

In chapter 6, a class of Lagrangian distribution, i.e. Lagrangian Hermite type probability distributions are discussed. The pmf and the cumulants of the basic Lagrangian Hermite distribution are studied. The parameters for the basic Lagrangian Hermite distribution are estimated by the method of moments and the method of first frequency and mean. For testing the validity of the estimate of the parameters of basic Lagrangian Hermite distribution the fitting of this distribution is considered. Then, several members of Lagrangian Hermite type distribution of type-I and type-II are investigated by various choice of pgf. The Lagrangian Hermite Poisson distribution of type-I and type-II are derived and fitted to some well known data sets with good results.

In chapter 7, a study on a class of Charlier type Lagrangian probability distribution has been made. The basic Lagrangian Charlier distribution is investigated. A discussion on some properties of pmf and the cumulant of the distribution are provided. The parameters are estimated by ratio of first two moments and first frequency. The fitting of the basic Lagrangian Charlier distribution has been considered for the testing of the validity of the estimates of the parameters. Further the general Lagrangian Charlier Poisson distribution of type-I and type-II are also investigated. An ad-hoc method is used for estimating the parameters of Lagrangian Charlier Poisson distribution of type-I. The fitting of the basic Lagrangian Charlier distribution is also compared with logarithmic series distribution, generalized logarithmic series distribution and geometric distribution and it is found that the basic Lagrangian Charlier distribution gives much closer fit than obtained by logarithmic series distribution, generalized logarithmic series distribution and geometric distribution.

CHAPTER 2

- RECURRENCE RELATIONS IN SOME DISCRETE INFLATED PROBABILITY DISTRIBUTIONS

Chapter 2

Recurrence Relations in some Discrete Inflated Probability Distributions

2.1 Introduction

Many distributions obtained in the course of experimental investigations often have an excess frequency of the observed event at zero point. This has been a major motivating force behind the development of inflated distributions that has been used as models in applied statistics. The inflated distribution is a finite mixture of original distribution. The probability mass function (pmf) of an inflated or zero modified distribution may be written as

$$P(X = x) = \begin{cases} \beta + \alpha p_0, & x = 0 \\ \alpha p_x, & x = 1,2,\ldots \end{cases} \tag{2.1.1}$$

where α is a parameter assuming arbitrary values in the interval (0,1] such that $\alpha+\beta=1$ and p_x is the pmf of the original distribution (for x=0,1,2,...). It is also possible to take the parameter β less than zero, provided $\beta + \alpha p_0 \geq 0$

$$\Rightarrow \qquad \beta \geq \frac{-p_0}{1-p_0} \tag{2.1.2}$$

Singh (1963) obtained the inflated Poisson distribution as a special case of contagious distribution. Further Singh (1965) indicated that there might exist analogous situations in binomial distribution, i.e. there is a distinct increase of the frequency of the observed event at zero point as well as a respective decrease of its value at the remaining points.

Pandey (1965) and Singh (1966) published the generalized distribution of equation (2.1.1) in the sense that the inflation of the distribution occurs at an arbitrary point 'l'. Thus the pmf of generalized inflated distribution is expressed as

$$P(X = x) = \begin{cases} \beta + \alpha p_l, & x = l \\ \alpha p_x, & x = 0,1,\ldots,l-1,l+1,\ldots \end{cases} \tag{2.1.3}$$

where $0 < \alpha \leq 1$ and $\alpha + \beta = 1$

Cohen (1960) gave some examples of fitting of inflated Poisson distributions to empirical data. Martin and Katti (1965) also fitted the distribution to a number of data sets. Khatri (1961) studied the logarithmic-with-zero distribution. Katti and Rao (1970) investigated log-zero-Poisson distribution and fitted this distribution by the method of maximum likelihood to each of the 35 empirical distributions collected by Martin and Katti (1965). Kemp and Kemp (1988) gave a bound for the maximum likelihood (ML) estimate of inflated Poisson distribution. They also studied a suitable method to provide initial ML estimators of the parameters of inflated binomial distribution.

The objective of this chapter is to extend the entire works on inflated distributions like binomial, Poisson, negative binomial and geometric etc. The essential problems considered in further examination of inflated distributions are -

(i) To obtain the recurrence relation for probabilities of certain inflated at 'l' distributions.

(ii) To calculate the moments recurrence relations for both inflated at zero and of inflated at 'l' distributions.

(iii) To obtain the rapid estimate of the parameters for inflated negative binomial and geometric distributions.

2.2 Inflated Binomial Distribution

The pmf of inflated at zero binomial distribution may be obtained from equation (2.1.1) as

$$P(X=x)=\begin{cases}\beta+\alpha q^{x}, & x=0\\ \alpha\binom{n}{x}p^{x}q^{n-x}, & x=1,2,3,\ldots\end{cases} \tag{2.2.1}$$

where $0<\alpha\leq 1$, $\beta=1-\alpha$, $0<p<1$ and $p+q=1$ [see Singh(1965)].

Similarly, for inflated at 'l' or for generalized inflated binomial distribution, the pmf may be written from equation (2.1.3) as

$$P(X=x)=\begin{cases}\beta+\alpha\binom{n}{x}p^{x}q^{n-x}, & x=l\\ \alpha\binom{n}{x}p^{x}q^{n-x}, & x=0,1,..l-1,l+1,..n\end{cases} \tag{2.2.2}$$

where $0< \alpha \leq 1$, $\beta=1-\alpha$, $0<p<1$ and $p+q=1$.

(a) Recurrence Relation for Probabilities

By using the relation of equation (2.2.1), the recurrence relation for probabilities may be derived as

$$P_{r+1} = \frac{(n-r)}{(r+1)}\frac{p}{q}P_r, \qquad r=1,2,3,....n \qquad (2.2.3)$$

where $P_r = \frac{\delta^r}{\delta t^r}H(t)\Big|_{t=0}$, $P_0 = \beta + \alpha q^n$, $P_1 = \alpha npq^{n-1}$ and $H(t)$ denotes the pgf of inflated at zero binomial distribution.

Similarly, for generalized inflated (inflated at 'l') binomial distribution the recurrence relation for probability may be derived as

$$P_r = \left\{\frac{(n+r-1)}{r}\right\}\frac{p}{q}P_{r-1}, \qquad r=1,2,...l-1,l+2,...n \qquad (2.2.4)$$

$$P_l = \beta + \alpha\binom{n}{l}p^l q^{n-l},\ r = l \quad \text{and} \quad P_{l+1} = \alpha\binom{n}{l+1}p^{l+1}q^{n-l-1},\ \text{for}\ r = l+1$$

where $P_0 = \alpha q^n$.

(b) Recurrence Relation for Raw Moments

The moment generating function (mgf) of a zero modified distribution may be easily derived from the original distribution. If M(t) is the mgf of the original distribution, then the mgf for inflated distribution may be written as

$$m(t) = \beta + \alpha M(t) \qquad \text{[see Johnson et al.(1992)]}$$

$$= \beta + \alpha\left(q + pe^{t}\right)^{n} \tag{2.2.5}$$

Differentiating both sides of equation (2.2.5) with respect to 't', we get

$$\left(q + pe^{t}\right)m'(t) = \alpha npe^{t}\left\{m(t) - \beta\right\} \tag{2.2.6}$$

Considering the coefficient of $\frac{t^r}{r!}$ in equation (2.2.6), we get

$$m'_{r+1} = p\left\{\alpha n\sum_{j=0}^{r}\binom{r}{j} - \sum_{j=0}^{r-1}\binom{r}{j+1}\right\}m'_{r-j} - \beta\alpha np \tag{2.2.7}$$

which is the recurrence relation for raw moments of inflated at zero binomial distribution where m'_r denotes the r[th] raw moment of inflated at zero binomial distribution. Putting r=0,1,2,3 in equation (2.2.7), the first four raw moments may be obtained as

$m'_1 = \alpha np$,

$m'_2 = \alpha n(n-1)p^2 + \alpha np$,

$m'_3 = \alpha n(n-1)(n-2)p^3 + 3\alpha n(n-1)p^2 + \alpha np$, and

$m'_4 = \alpha n(n-1)(n-2)(n-3)p^4 + 6\alpha n(n-1)(n-2)p^3 + 7\alpha n(n-1)p^2 + \alpha np$

respectively.

Hence mean= $m'_1 = \alpha np$ and variance= $\alpha np(1 - p + \beta np)$. (2.2.8)

To derive the recurrence relation for the simple moments of inflated at 'l' binomial distribution, we consider the lemma used by Gerstenkorn (1979). According to his lemma

$$\phi_y(t) = \beta e^{ilt} + \alpha\phi_x(t) \tag{2.2.9}$$

where $\phi_y(t)$ and $\phi_x(t)$ denote the characteristic functions of a random variable Y and X , with and without inflation respectively.

The characteristics function of the random variable possessing the moments of an arbitrary order can be expanded by using the Maclaurian series as

$$\phi_y(t) = \sum_{j=0}^{\infty} \frac{(it)^j}{j!} m_j \tag{2.2.10}$$

where m_j denotes the j[th] simple moments of the random variable Y, which follows the inflated distribution. Then putting $it = \theta$ in equation (2.2.10), the following relation may be obtained

$$\sum_{j=0}^{\infty} \frac{\theta^j}{j!} m_j = \beta e^{\alpha} + \alpha\left(q + pe^{\theta}\right)^n \tag{2.2.11}$$

Differentiating both sides of equation (2.2.11) with respect to θ, we get

$$\left(q + pe^{\theta}\right)\sum_{j=1}^{\infty} \frac{\theta^{j-1}}{(j-1)!} m_j = \left(q + pe^{\theta}\right)\alpha l e^{\alpha} + \alpha n p e^{\theta}\left(q + pe^{\theta}\right)^n \tag{2.2.12}$$

Considering the following transformation in the left hand side of equation (2.2.11), we get

$$\begin{aligned}\left(q + pe^{\theta}\right) &= p\left(e^{\theta} - 1\right) + q + p \\ &= p\left(e^{\theta} - 1\right) + 1 \\ &= p\sum_{i=1}^{\infty} \frac{\theta^i}{i!} + 1\end{aligned}$$

$$\Rightarrow \sum_{j=1}^{\infty} \frac{\theta^{j-1}}{(j-1)!} m_j = \left(q + pe^{\theta}\right)\beta l e^{\alpha} + npe^{\theta}\left(\sum_{j=0}^{\infty} \frac{\theta^j}{j!} m_j - \beta e^{\alpha}\right) - p\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\theta^{i+j-1}}{(j-1)!} m_j$$

$$= q\beta l\sum_{i=0}^{\infty} \frac{\theta^i l^i}{i!} + p\beta(l-n)\sum_{i=0}^{\infty} \frac{\theta^i (l+1)^i}{i!} + np\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\theta^{i+j}}{i!\,j!} m_j - p\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\theta^{i+j-1}}{(j-1)!} m_j$$

Considering the coefficients of θ^{r-1}, we get

$$m_r = \beta q l^r + \beta p(l-n)(l+1)^{r-1} + p\sum_{k=0}^{r-1} n\left[\binom{r-1}{k} - \binom{r-1}{k+1}\right] m_{r-k-1} \tag{2.2.13}$$

which is the recurrence relation for raw moments of inflated at '*l*' binomial distribution where m_r denotes the r[th] raw moment of the distribution. Considering r=1,2,3,4 in equation (2.2.13), we get

$m_1 = \beta l + \alpha np$,

$m_2 = \beta l^2 + \alpha n(n-1)p^2 + \alpha np$,

$m_3 = \beta l^3 + \alpha n(n-1)(n-2)p^3 + 3\alpha n(n-1)p^2 + \alpha np$,

$m_4 = \beta l^4 + \alpha n(n-1)(n-2)(n-3)p^4 + 6\alpha n(n-1)(n-2)p^3 + 7\alpha n(n-1)p^2 + \alpha np$

respectively, where m_r denotes the r[th] raw moments for inflated at '*l*' binomial distribution. Thus the above raw moments provide an illustration of the simple formula

$$m_r = \beta l^r + \alpha m'_r \tag{2.2.14}$$

where m'_r is the r[th] raw moments of when there is no inflation. Hence equation (2.2.14) determines the relationship between the raw moments of the inflated distribution and the one without inflation.

(c) Recurrence Relation for Central moments

From the definition of the central moment of r[th] order of generalized inflated binomial distribution, we have

$$\mu_r = \beta(l - m_1)^r + \alpha\sum_{x=0}^{n}(x - m_1)^r \binom{n}{x} p^x q^{n-x} \tag{2.2.15}$$

Putting the values of $m_1 = \beta l + \alpha np$ in equation (2.2.15), we may obtain

$$\mu_r = \beta(l - \beta l - \alpha np)^r + \alpha\sum_{x=0}^{n}(x - \beta l - \alpha np)^r \binom{n}{x} p^x q^{n-x} \tag{2.2.16}$$

Differentiating both sides of equation (2.2.16) with respect to 'p', we get

$$\frac{\delta\mu_r}{\delta p} = -\alpha^r \beta nr(l - np)^{r-1} - \alpha^2 nr\sum_{x=0}^{n}(x - l\beta - \alpha np)^r \binom{n}{x} p^x q^{n=x} + \frac{\alpha}{pq}\sum_{x=0}^{n}\binom{n}{x} p^x$$

$$q^{n-x}(x - l\beta - \alpha np)^r x - \frac{\alpha n}{q}\sum_{x=0}^{n}\binom{n}{x} p^x q^{n-x}(x - l\beta - \alpha np)^r \tag{2.2.17}$$

Substituting x=$(x - l\beta - \alpha np) + (l\beta + \alpha np)$ in equation (2.2.17), we obtain the following relation after simple transformation.

$$\frac{\delta\mu_r}{\delta p} = -\alpha nr\mu_{r-1} + \frac{\alpha}{pq}\sum_{x=0}^{n}(x - l\beta - \alpha np)^{r+1}\binom{n}{x} p^x q^{n-x} + \frac{\alpha\beta(l - np)}{pq}$$

$$\sum_{x=0}^{n}(x - l\beta - \alpha np)^r \binom{n}{x} p^x q^{n-x} \tag{2.2.18}$$

It follows from equation (2.2.16) that

$$\alpha\sum_{x=0}^{n}(x - l\beta - \alpha np)^t \binom{n}{x} p^x q^{n-x} = \mu_t - \alpha^t \beta(l - np)^t$$

Thus from equation (2.2.18), we have

$$\frac{\delta\mu_r}{\delta p} = -\alpha nr\mu_{r-1} + \frac{1}{pq}\left\{\mu_{r+1} - \alpha^{r+1}\beta(l - np)^{r+1}\right\} + \frac{\beta(l - np)}{pq}\left\{\mu_r \alpha^r \beta(l - np)^{r+1}\right\}$$

$$\Rightarrow \quad \mu_{r+1} = pq\frac{\delta\mu_r}{\delta p} + \frac{\alpha nr}{pq}\mu_{r-1} - \beta(l-np)\mu_r + \alpha^r\beta(l-np)^{r+1} \qquad (2.2.19)$$

Putting $l=0$ in equation (2.2.19), we obtain the recurrence relation for central moments of inflated at zero binomial distribution. The equation (2.2.19) with $\beta = 0$ and $l = 0$ gives the recurrence relation for the central moments of the binomial distribution without inflation, i.e.

$$\mu_{r+1} = pq\left(\frac{\delta\mu_r}{\delta p} + nr\mu_{r-1}\right) \qquad (2.2.20)$$

2.3 Inflated Poisson Distribution

The pmf of inflated Poisson distribution at zero may be written as

$$P(X=x) = \begin{cases} \beta + \alpha e^{-\phi}, & x = o \\ \alpha\frac{e^{-\phi}\phi^x}{x!}, & x = 1,2,\ldots \end{cases} \qquad (2.3.1)$$

and for inflated at 'l', the pmf may be written as

$$P(X=x) = \begin{cases} \beta + \alpha\frac{e^{-\phi}\phi^x}{x!}, & x = l \\ \alpha\frac{e^{-\phi}\phi^x}{x!}, & x = 0,1,..l-1,l+1,\ldots \end{cases} \qquad (2.3.2)$$

where $0<\alpha \le 1$, $\alpha + \beta = 1$ and $\phi > 0$

(a) Recurrence Relation for Probabilities

Recurrence relation for probabilities of inflated at zero Poisson may be written as

$$P_{r+1} = \frac{\phi}{r+1}P_r, \qquad r=1,\ 2,\ldots, \qquad (2.3.3)$$

where $P_0 = \beta + \alpha e^{-\phi}$, $P_1 = \alpha\phi e^{-\phi}$.

Similarly, for inflated at 'l' distribution, the recurrence relation for probability may be written as

$$P_r = \frac{\phi}{r} P_{r-1}, \qquad r=1, 2, \ldots l-1, l+2, \ldots\ldots \tag{2.3.4}$$

where $P_l = \beta + \alpha \frac{e^{-\phi}\phi^l}{l!}$, $P_{l+1} = \alpha \frac{e^{-\phi}\phi^{l+1}}{(l+1)!}$ and $P_0 = \alpha e^{-\phi}$

(b) Recurrence Relation for Raw Moments

If $m(t)$ is mgf of inflated Poisson distribution then, we have

$$m(t) = \beta + \alpha e^{\phi(e^t - 1)} \tag{2.3.5}$$

Differentiating both sides of equation (2.3.5) with respect to 't', and equating the term $\frac{t^r}{r!}$, we get the raw moment recurrence relation as

$$m'_{r+1} = \sum_{j=0}^{r} \binom{r}{j} m'_{r-j} - \beta , \quad r=0,1,2,\ldots \tag{2.3.6}$$

where m'_r denotes the r[th] order raw moments for inflated Poisson distribution. Putting $r = 0,1,2,3$ in equation (2.3.6) the first four raw moments of inflated Poisson distribution may be obtained as

$$m'_1 = \alpha\phi ,$$

$$m'_2 = \alpha\phi(\phi + 1),$$

$$m'_3 = \alpha\phi(\phi^2 + 3\phi + 1),$$

and $\quad m_4' = \alpha\phi\left(\phi^3 + 6\phi^2 + 7\phi + 1\right)$

respectively. Hence mean= $\alpha\phi$, variance= $\alpha\phi(1 + \phi\beta)$.

Similarly, if $\phi_x(t)$ and $\phi_y(t)$ denote characteristics function of Poisson and generalized inflated Poisson distribution respectively, then using Gerstenkorn (1979) lemma, we have

$$\phi_y(t) = \beta e^{itl} + \alpha\phi_x(t) \tag{2.3.7}$$

After some suitable transformations, the recurrence relation for raw moments of generalized inflated Poisson distribution may be written as

$$m_{r+1} = \beta l^{r+1} - \beta\phi(l+1)^r + \phi\sum_{j=0}^{r}\binom{r}{j}m_{r-j} \qquad r=0,1,2,\ldots \tag{2.3.8}$$

Putting r=0,1,2,3 in equation (2.3.8), we get

$$m_1 = \beta l + \alpha\phi\,,$$

$$m_2 = \beta l^2 + \alpha\phi(\phi + 1),$$

$$m_3 = \beta l^3 + \alpha\phi\left(\phi^2 + 3\phi + 1\right),$$

and $\quad m_4 = \beta l^4 + \alpha\phi\left(\phi^3 + 6\phi''^2 + 7\phi + 1\right)$

respectively, where m_r denotes the rth raw moment of inflated at 'l' Poisson distribution.

2.4 Inflated Negative Binomial Distribution

The pmf of inflated negative binomial distribution may be written from equation (2.1.1) as

$$P(X=x)=\begin{cases}\beta+\alpha p^{n}, & x=0\\ \alpha(-1)^{x}\binom{-n}{x}p^{n}q^{x}, & x=1,2,\ldots\end{cases} \tag{2.4.1}$$

where $0<\alpha\le 1$, $\alpha+\beta=1$, $0<p<1$ and $p+q=1$

Similarly, from equation (2.1.3), the pmf of generalized inflated negative binomial distribution may be written as

$$P(X=x)=\begin{cases}\alpha(-1)^{x}\binom{-n}{x}p^{n}q^{x}, & x=0,1,2,\ldots,l-1\\ \beta+\alpha(-1)^{x}p^{n}q^{l}, & x=l\\ \alpha(-1)^{x}\binom{-n}{x}p^{n}q^{x}, & x=l+1,l+2,\ldots\end{cases} \tag{2.4.2}$$

where $0<\alpha\le 1$, $\alpha+\beta=1$, $0<p<1$ and $p+q=1$

(a) Recurrence Relation for Probabilities

The pgf of inflated negative binomial distribution may be written as

$$H(t)=\beta+\alpha p^{n}(1-qt)^{-n} \tag{2.4.3}$$

Differentiating both sides of equation (2.4.3) with respect to 't', and equating the coefficient of t^{r}, the recurrence relation for probabilities may be obtained as

$$P_{r+1}=\frac{(n+r)q}{(r+1)}P_{r},\ \text{r=1,2,...} \tag{2.4.4}$$

where $P_0=\beta+\alpha p^{n}$ and $P_1=\alpha n p^{n}q$

(b) Recurrence Relation for Raw Moments

The mgf for inflated negative binomial distribution may be written as

$$m(t)=\beta+\alpha p^{n}\left(1-qe^{t}\right)^{-n} \tag{2.4.5}$$

Differentiating both sides of equation (2.4 5) with respect to 't' and equating the coefficient of $t^r/r!$, and after some suitable calculations it is obtained that

$$m'_{r+1} = \frac{1}{p}\left[q\left\{n\sum_{j=0}^{r}\binom{r}{j} + \sum_{j=0}^{r-1}\binom{r}{j+1}\right\}m'_{r-j} - \beta np\right] \qquad (2.4.6)$$

Putting r=0,1,2,3 in equation (2.4.6), the first four raw moments may be written as follows

$$m'_1 = \alpha n\frac{q}{p},$$

$$m'_2 = \alpha\left\{n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\},$$

$$m'_3 = \alpha\left\{n(n+1)(n+2)\frac{q^3}{p^3} + n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\}, \text{ and}$$

$$m'_4 = \alpha\left\{n(n+1)(n+2)(n+3)\frac{q^4}{p^4} + n(n+1)(n+2)\frac{q^3}{p^3} + n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\}$$

respectively, where m'_r denotes the r^th raw moments of inflated at zero negative binomial distribution.

The recurrence relation for moments of generalized inflated negative binomial distribution may be written as

$$m_r = \frac{1}{p}\left[\beta l^2 - \beta q(l+n)(l+1)^{r-1} + q\sum_{j=0}^{r-1}\left\{n\binom{r-1}{j} + \binom{r-1}{j+1}\right\}m_{r-j-1}\right] \qquad (2.4.7)$$

Thus the first four raw moments of generalized inflated negative binomial distribution may be obtained by putting r=1,2,3,4 in equation (2.4.7) as

$$m_1 = \beta l + \alpha n\frac{q}{p},$$

$$m_2 = \beta l^2 + \alpha\left\{n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\},$$

$$m_3 = \beta l^3 + \alpha\left\{n(n+1)(n+2)\frac{q^3}{p^3} + n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\},\ \text{and}$$

$$m_4 = \beta l^4 + \alpha\left\{n(n+1)(n+2)(n+3)\frac{q^4}{p^4} + n(n+1)(n+2)\frac{q^3}{p^3} + n(n+1)\frac{q^2}{p^2} + n\frac{q}{p}\right\}.$$

respectively, where m_r denotes the r^{th} raw moments for generalized inflated negative binomial distribution.

2.5 Inflated Geometric Distribution

The pmf for inflated at zero geometric distribution may be written as

$$P(X = x) = \begin{cases} \beta + \alpha p, & x = 0 \\ \alpha p q^x, & x = 1,2,3\ldots \end{cases} \tag{2.5.1}$$

where $0 < \alpha \le 1$, $\alpha + \beta = 1$, $0 < p < 1$ and $p + q = 1$

Similarly, for inflated at 'l' geometric distribution, the pmf may be written as

$$P(X = x) = \begin{cases} \alpha p q^x, & x = 0,1,..l-1 \\ \beta + \alpha p q^x, & x = l \\ \alpha p q^x, & x = l+1, l+2,\ldots \end{cases} \tag{2.5.2}$$

where, $0 < \alpha \le 1$, $\alpha + \beta = 1$, $0 < p < 1$ and $p + q = 1$.

(a) Recurrence Relation for Probabilities

Since geometric distribution is a special case of negative binomial distribution, so putting n=1 in (2.4.4), we get the following recurrence relation for probabilities of inflated geometric distribution as

$$P_{r+1} = qP_r, \qquad r=1,2,3,\ldots \tag{2.5.3}$$

where $P_0 = \beta + \alpha p$ and $P_1 = \alpha pq$

Similarly, for inflated at '*l*' geometric distribution, the recurrence relation for probabilities may be written as

$$P_r = \begin{cases} \alpha p, & r=0 \\ qP_{r-1}, & r=1,2,\ldots,l-1 \\ \beta + \alpha pq^l, & r=l \\ \alpha pq^r, & r=l+1,l+2,\ldots \end{cases} \tag{2.5.4}$$

(b) Recurrence Relation for Raw Moments

The raw moment recurrence relation of inflated geometric distribution may be obtained from equation (2.4.6) by putting n=1.

$$m'_{r+1} = \frac{q}{p}\left\{\sum_{j=0}^{r}\left(\frac{r!}{(r-j)!\,j!}\right) + \sum_{j=0}^{r-1}\left(\frac{r!}{(j+1)!\,(r-j-1)!}\right)\right\} m'_{r-j} - \frac{q}{p}\beta \tag{2.5.6}$$

Similarly, putting n=1 in equation (2.4.7) of generalized inflated negative binomial distribution, the recurrence relation for generalized inflated geometric distribution may be obtained as

$$m_{r+1} = \frac{1}{p}\left[\beta l^r - q\beta(l+1)^r + q\sum_{k=0}^{r-1}\left\{\binom{r-1}{k} + \binom{r-1}{k+1}\right\} m_{r-1-k}\right] \tag{2.5.7}$$

Putting r=0,1,2,3 respectively in equation (2.5.7), the first four raw moments for the generalized inflated geometric distribution may be obtained as

$$m_1 = \beta l + \alpha \frac{q}{p},$$

$$m_2 = \beta l^2 + \alpha \left\{ 2\frac{q^2}{p^2} + \frac{q}{p} \right\},$$

$$m_3 = \beta l^3 + \alpha \left\{ 6\frac{q^3}{p^3} + 6\frac{q^2}{p^2} + \frac{q}{p} \right\},$$

and $$m_4 = \beta l^4 + \alpha \left\{ 24\frac{q^4}{p^4} + 36\frac{q^3}{p^3} + 14\frac{q^2}{p^2} + \frac{q}{p} \right\}$$

respectively, where m_r denotes the rth raw moments for generalized inflated geometric distribution. Considering *l=0* in the above relations, the first four raw moments of inflated at zero geometric distribution may be obtained.

2.6 Inflated Power Series Distribution

Patel (1975) introduced the inflated at zero power series distribution. He showed that Singh's inflated Poisson distribution is a particular case of inflated power series distribution.

The pmf of inflated at zero power series distribution may be written from equation (2.1.1) as

$$P(X = x) = \begin{cases} \beta + \alpha \frac{a_0}{f(\theta)}, & x = 0 \\ \alpha \frac{a_x \theta^x}{f(\theta)}, & x = 1,2,\ldots \end{cases} \qquad (2.6.1)$$

where $0 < \alpha \le 1, \beta = 1 - \alpha,\ f(\theta) = \sum_x a_x \theta^x$, $a_j \ge 0$ and $\theta > 0$.

In (2.6.1), θ is the power parameter and $f(\theta)$ is the series function.

Similarly, the pmf of inflated at 'l' power series distribution may be written from equation (2.1.3) as

$$P(X=x)=\begin{cases}\beta+\alpha\dfrac{a_x\theta^x}{f(\theta)}, & x=l\\ \alpha\dfrac{a_x\theta^x}{f(\theta)}, & x=0,1,2,\ldots,l-1,l+1,\ldots\end{cases} \tag{2.6.2}$$

Grzegorska (1973) studied the inflated generalized power series distribution (IGPD). He obtained the moment recurrence relation of IGPD.

(a) Recurrence Relation for factorial Moments

The factorial moment generating function (fmgf) of inflated power series distribution may be written as

$$H(1+t)=\beta+\alpha\frac{f(\theta+t\theta)}{f(\theta)} \tag{2.6.3}$$

So the rth factorial moments of inflated power series distribution may be written as

$$\mu'_{(r)}=\left[\frac{d^r}{dt^r}H(1+t)\right]_{t=0}=\frac{\alpha\theta^r}{f(\theta)}\frac{d^r}{d\theta^r}[f(\theta)] \tag{2.6.4}$$

Putting r=1, 2 in equation (2.6.4), we get the first two factorial moments as

$$\mu'_{(1)}=\frac{\alpha\theta}{f(\theta)}f'(\theta),$$

$$\mu'_{(2)}=\frac{\alpha\theta}{f(\theta)}f''(\theta),$$

respectively, where $\mu'_{(r)}$ denotes the rth factorial moments of inflated power series distribution. Hence the mean and variance of inflated power series distribution are

$$\text{Mean} = \frac{\alpha\theta}{f(\theta)} f'(\theta) \quad \& \quad \text{Variance} = \frac{\alpha\theta^2}{f(\theta)} f''(\theta) + \frac{\alpha\theta}{f(\theta)} f'(\theta) - \frac{\alpha^2\theta^2}{[f(\theta)]^2}[f'(\theta)]^2.$$

In particular, considering the different values of θ and $f(\theta)$ in equation (2.6.4), we can find the mean and variance of inflated binomial, inflated Poisson and inflated negative binomial distribution.

(b) Recurrence Relation for Central Moments

From the definition of the central moments, the rth central moment of inflated at '*l*' power series distribution may be written as

$$\mu_r = \beta(l - m_1)^r + \alpha \sum_x (x - m_1)^r \frac{a_x \theta^x}{f(\theta)} \tag{2.6.5}$$

where $m_1 = l\beta + \alpha \frac{\theta f'(\theta)}{f(\theta)}$ is the mean of inflated at '*l*' power series distribution.

Differentiating both sides of equation (2.6.5) with respect to θ and after some suitable calculation, the recurrence relation for central moments may be written as

$$\mu_{r+1} = \theta\left[\frac{\delta\mu_r}{\delta\theta} + r\frac{\delta m_1}{\delta\theta}\mu_{r-1}\right] - \frac{\beta}{\alpha}(l - m_1)\mu_r + \frac{\beta}{\alpha}(l - m_1)^{r+1} \tag{2.6.6}$$

Putting $l=0$ in equation (2.6.6), we obtain the recurrence relation for central moments of inflated at zero power series distribution. After some simple transformation of equation (2.6.6), we can easily derive the recurrence relation for central moment of inflated at '*l*' binomial distribution as shown in equation (2.2.19). Similarly, the recurrence relation for central moments of inflated

Poisson, negative binomial and geometric distribution can be derived from equation (2.6.6).

(c) Some Special Cases of Inflated Power Series Distribution.

Noack (1950) first studied the power series distribution and showed that many important discrete distributions like binomial, Poisson, negative binomial and logarithmic belong to this class. Similarly the limiting forms of the inflated power series distribution are given in Table 2.1 as the parameter takes different values.

Table 2.1

Sl.No.	Parameter values	Distribution and its pgf
1	$f(\theta)=(1+\theta)^n$, n is +ve integer $\theta=\frac{p}{(1-p)}$, $a_x=\binom{n}{x}$	Inflated binomial, $\beta+\alpha(q+pt)^n$
2	$f(\theta)=e^{\phi}$, $\theta=\phi$, $a_x=\frac{1}{x!}$	Inflated Poisson, $\beta+\alpha e^{\phi(t-1)}$
3	$f(\theta)=(1-\theta)^{-n}$, k>0 $\theta=q$, $a_x=(-1)^x\binom{-n}{x}$	Inflated negative binomial, $\beta+\alpha p^n(1-qt)^{-n}$

2.7 Some Properties of Inflated Distribution.

(a) Distribution of the sum.

Sobich and Szynal (1974) studied the distribution of the sum of 'm' independent random variables having the same inflated binomial distribution. They used the characteristics function to obtain the distribution of the sums.

Lingappaiah (1977) also obtained the distribution of sum of 'm' independent random variables having the same inflated distribution. Let $\phi_0(t)$ be the characteristic function of the inflated at zero distribution while $\phi(t)$ be the characteristic function of non inflated case. Then we have

$$\phi_0(t) = \beta + \alpha\phi(t) \tag{2.7.1}$$

If $X_1, X_2, \ldots, X_m$ are the independent and identical random variables having the same inflated distribution as in equation (2.1.1) and if $Z = X_1 + X_2 + \ldots + X_m$, then the characteristic function of the sum of 'm' independent and identical random variables having the same inflated power series distribution may be written as

$$\phi_0^m(t) = [\beta + \alpha\phi(t)]^m \tag{2.7.2}$$

Using the Inversion form for characteristic function we have the distribution of the sum of 'm' independent random variables having the same inflated power series distribution as

$$P(Z = z) = \begin{cases} (\beta + \alpha P_0)^m, & z = 0 \\ \sum_{r=1}^{m} \binom{m}{r} \beta^{m-r} \alpha^r P_{s_r}(z), & z = 1,2,\ldots \end{cases} \tag{2.7.3}$$

where $P_{s_r}(z)$ is the distribution of $X_1 + X_2 + \ldots + X_r$, and $P_0 = \frac{a_0}{f(\theta)}$.

(b) Truncated Case

The distribution of truncated inflated distribution, truncated at zero, is same as the truncated simple distribution truncated at the same point [see Sobich and Szynal, (1974)]. If $P_T(X = x)$, denotes the probability function of truncated inflated distribution then

$$P_T(X = x) = \frac{\alpha P_x}{1-[\beta+\alpha P_0]} = \frac{P_x}{1-P_0}, \qquad \text{for } x=1,2,3,\ldots \qquad (2.7.7)$$

where $P_x = P(X = x)$, for x=1,2,3,... and $P_0 = P(X = 0)$.

2.8 Estimation of Parameters of Inflated distributions

The estimation of parameters of inflated distributions other than α can be carried out by ignoring the observed frequency in the zero class and then using a technique appropriate to the original distribution. After the other parameters have been estimated, the value of α can then be estimated by equating the observed and expected frequencies in the zero class [see Johnson et al.(1992)].

(a) Inflated Binomial Distribution

(i) Method of maximum likelihood

Since this is a zero modified distribution, hence, one of the maximum likelihood (ML) equation is [see Johnson et al., (1992), p. 315]

$$\hat{\beta} + (1-\hat{\alpha})q^n = \frac{f_0}{N} \qquad (2.8.1)$$

where $\frac{f_0}{N}$ is the observed proportion of zeros. As it is a power series distribution therefore, the other ML equation is

$$\bar{x} = n\hat{\alpha}\hat{p} \qquad (2\ 8.2)$$

Kemp and Kemp (1988) showed that these maximum likelihood equations do not have explicit solution, therefore, for rapid assessment of data they used method of moments and method of mean and first frequency.

(ii) Method of moments

By equating the sample mean $\bar{x}$ and sample variance s^2 to the population mean and variance respectively in equation (2.2.8), we get

$$\tilde{p} = \frac{\frac{s^2}{\bar{x}} + \bar{x} - 1}{n-1} \tag{2.8.3}$$

and

$$\tilde{\alpha} = \frac{\bar{x}}{n\tilde{p}} \tag{2.8.4}$$

(iii) Method of mean and first frequency

By equating the 1st class probability with $\frac{f_1}{N}$, we obtain

$$p' = 1 - \left\{\frac{f_1}{Nx}\right\}^{\frac{1}{(n-1)}} \tag{2.8.5}$$

and

$$\alpha' = \frac{\bar{x}}{np'} \tag{2.8.6}$$

where $\frac{f_1}{N} = n\alpha p q^{n-1}$.

(b)Inflated Poisson Distribution

(i)Method of maximum likelihood

The ML equations for inflated Poisson distribution are [Singh, (1963)]

$$\hat{\beta} + \hat{\alpha} e^{-\hat{\phi}} = \frac{f_0}{N} \quad (2.8.7)$$

$$\bar{x} = \hat{\alpha}\hat{\phi} \quad (2.6.8)$$

Eliminating $\hat{\alpha}$, from equations (2.8.7) and (2.8.8) gives

$$\bar{x}\left(1 - e^{-\hat{\phi}}\right) = \hat{\phi}\left(1 - \frac{f_0}{N}\right) \quad (2.8.9)$$

Hence $\hat{\phi}$ (and $\hat{\alpha}$) can be obtained by iteration, [see Johnson et al. (1992), p. 314]. Martin and Katti (1965) fitted this distribution to 35 data sets, using ML method with $\phi^{(0)} = \dfrac{\bar{x}}{\left(1 - f_0/N\right)}$ as the initial estimate of ϕ. Kemp (1986) showed that $\phi_{(0)} = In\left(\dfrac{N\bar{x}}{f_1}\right)$ was another initial estimate of ML equation and it was found that usually $\phi_{(0)} < \hat{\phi} < \phi^{(0)}$.

(ii) Method of moments

By equating sample mean $\bar{x}$ and sample variance s^2 to the population mean and variance of inflated Poisson distribution, the following relations are obtained.

$$\bar{x} = \tilde{\alpha}\tilde{\phi} \quad (2.8.10)$$

and $$s^2 = \tilde{\alpha}\tilde{\phi}'\{1 + \tilde{\phi}(1 - \tilde{\alpha})\} \quad (2.8.11)$$

$$\Rightarrow \tilde{\phi} = \frac{s^2}{\bar{x}} + \bar{x} - 1 \quad (2.8.12)$$

and $$\tilde{\alpha} = \frac{\bar{x}}{\tilde{\phi}} \tag{2.8.13}$$

(iii) Method of mean and first frequency

By equating the 1st probability with $\frac{f_1}{N}$, where

$$\frac{f_1}{N} = \alpha' e^{-\phi'} \tag{2.8.14}$$

and $$\bar{x} = \alpha' \phi' \tag{2.8.15}$$

$$\Rightarrow \quad \phi' = \log\left(\frac{N\bar{x}}{f_1}\right) \tag{2.8.16}$$

and $$\alpha' = \frac{\bar{x}}{\phi'} \tag{2.8.17}$$

(c) Inflated Negative Binomial Distribu

A composite method has been used to estimate the parameters of inflated negative binomial (INB) distribution.

(i) Ratio of first two moments and ratio of frequencies

By equating the sample mean $\bar{x}$ and sample variance s^2 to the corresponding population values, we have

$$\bar{x} = \alpha n \frac{q}{p} \tag{2.8.18}$$

$$s^2 = \alpha n(n+1)\frac{q^2}{p^2} + \alpha n \frac{q}{p} - \alpha^2 n^2 \frac{q^2}{p^2} \tag{2.8.19}$$

$$\Rightarrow \quad \frac{s^2}{\bar{x}} + \bar{x} - 1 = (n+1)\frac{q}{p} \tag{2.8.20}$$

Again, we have $\frac{f_2}{N} = \frac{(n+1)q}{2}\frac{f_1}{N}$

$$\Rightarrow \quad \frac{f_2}{f_1} = \frac{(n+1)q}{2} \tag{2.8.21}$$

From equation (2.8.20) and (2.8.21), we may have

$$\hat{p} = \frac{\frac{2f_2}{f_2}}{\frac{s^2}{\bar{x}} + \bar{x} - 1} \tag{2.8.22}$$

From equation (2.8.22) and the first frequency $\frac{f_1}{N} = n\alpha p^n q$, we get

$$\hat{n} = 1 + \frac{\log(f_1 / N\bar{x})}{\log \hat{p}} \tag{2.8.23}$$

Again from the zero class frequency $\frac{f_0}{N} = \beta + \alpha q^n$, we get

$$\hat{\alpha} = \frac{1 - \frac{f_0}{N}}{1 - \hat{p}^{\hat{n}}} \tag{2.8.24}$$

(d)Inflated Geometric Distribution

The parameters of inflated geometric distribution may be estimated by following method of estimation.

(i) Method of Maximum Likelihood (ML)

Since the inflated geometric distribution is a zero modified distribution so one of the ML equation is

$$\hat{\beta} + \hat{\alpha}\hat{p} = \frac{f_0}{N} \tag{2.8.25}$$

where $\frac{f_0}{N}$ is the observed proportion of zeros. It is also a power series distribution, and so the other ML equation is

$$\bar{x} = \frac{\hat{q}}{\hat{p}}\hat{\alpha} \tag{2.8.26}$$

Eliminating $\hat{\alpha}$ between equation (2.8.25) and (2.8.26), we get

$$\hat{p} = \frac{\left(1 - \frac{f_0}{N}\right)}{\bar{x}} \tag{2.8.27}$$

and $$\hat{\alpha} = \frac{\left(1 - \frac{f_0}{N}\right)}{\hat{q}} \tag{2.8.28}$$

(ii) Method of moments.

The parameters p and α of inflated at zero geometric distribution may be estimated from the mean and variance as

$$\tilde{p} = \frac{2\bar{x}}{s^2 + \bar{x}^2 + \bar{x}} \tag{2.8.29}$$

and $$\tilde{\alpha} = \frac{\bar{x}\tilde{p}}{(1 - \tilde{p})} \tag{2.8.30}$$

where mean $= \bar{x} = \tilde{\alpha}\frac{\tilde{q}}{\tilde{p}}$ and variance$= s^2 = \tilde{\alpha}\frac{\tilde{q}}{\tilde{p}}\left\{\frac{\tilde{q}}{\tilde{p}}(2 - \tilde{\alpha}) + 1\right\}$

(iii) Method of mean and first frequency

We have $\frac{f_1}{N} = \tilde{\alpha}\tilde{p}\tilde{q}$ and $\bar{x} = \frac{\tilde{\alpha}\tilde{q}}{\tilde{p}}$

$$\Rightarrow \quad \tilde{p} = \sqrt{\frac{f_1}{N\bar{x}}} \tag{2.8.31}$$

and $\quad \tilde{\alpha} = \dfrac{\widetilde{xp}}{\tilde{q}}$ (2.8.32)

2.9 Goodness of fit

Here we consider some reported data sets for fitting of these inflated binomial, Poisson and geometric distribution. We use the method of maximum likelihood, method of moments, method of mean and first frequency for fitting of these distributions.

In Table 2.2, the inflated binomial distribution was fitted to Pyrausta ni bilalis data [Beall (1940)], for which Neyman type A (NTA) and Neyman type B (NTB) were fitted by McGuire et al (1957). Since maximum likelihood method does not give an explicit solution, method of moment and an adhoc method are used here to estimate the parameters of the inflated binomial distribution. It has been observed that when the first two frequencies are large compared to others, the estimation by the adhoc method gives some improvement in the fitting of the inflated binomial distribution.

In Table 2.3 and 2.4, we consider European corn borer data for which negative binomial and Neyman type A distributions were fitted by McGuire et al (1957) and inflated Poisson distribution (ML method) was fitted by Singh (1963). It is observed from the tables (2.3 & 2.4) that method of moments, method of mean and first frequency give as good fit as the maximum likelihood method by Singh (1963). Thus in most of the cases it has been seen that if there is an excess frequency of observed event at point zero as well as a respective decrease of its value at the remaining points, the inflated Poisson distribution provides much closer fit to the data than the others.

In Table 2.5, for fitting of inflated geometric distribution, the accident of 647 women working in a high explosive shells during 5 weeks period [data Greenwood & Yule (1920)] has been considered for which negative binomial and Hermite distributions were fitted by Plunket & Jain (1975). In Table 2.6, European corn borer data for which negative binomial distribution (ML method) was fitted by McGuire et al (1957) has been considered for fitting of inflated geometric distribution. It is observed from the Table 2.5 & 2.6, that maximum likelihood method gives better fit in both the cases as judges by the χ^2 values.

Here the fitting of inflated negative binomial distribution is not considered since the fitting of this distribution by the composite method does not give better fit like inflated Poisson, inflated binomial and inflated geometric distribution. Hence the result is not reported in this case.

Table 2.2 Observed and expected frequencies on the basis of inflated binomial (IB), NTA & NTB distribution. [Data on Pyrausta nibilalis to which NTA & NTB were fitted by Beall and Rescia (1953)]

No of insects	Observed frequency	Expected frequency			
		IB(MM)	IB(FM)	NTA	NTB
0	33	34.35	32.00	37.8	37.1
1	12	6.66	11.17	5.6	6.8
2	6	7.79	8.21	5.2	5.0
3	3	4.55	3.02	3.5	3.2
4	1	1.33	0.56	1.9	1.9
≥ 5	1	0.32	0.04	2.0	2.0
Total	56	56.00	56.00	56.0	56.0
χ^2		5.0396	0.1083	9.04	5.57
Parameter estimates $\hat{p}$		0.3690	0.2689		
$\hat{\alpha}$		0.4138	0.7311		

Note: IB: Inflated Binomial distribution,

NTA: Neyman Type A,

NTB: Neyman Type B,

MM: Method of moments,

FM: Method of Mean and first frequency.

Table 2.3 Observed and expected frequencies on the basis IP, NB and NTA distribution. [Data McGuire et al (1957), distribution 6]

Count per plant	Observed frequency	Expected frequency				
		IP(MM)	IP(FM)	IP(ML)	NB	NTA
0	907	906.19	906.53	907.00	902.85	906.28
1	275	275.54	275.94	274.27	288.86	280.61
2	88	90.64	90.73	90.85	78.07	82.01
3	23	19.88	19.96	20.06	19.81	20.45
4	3	3.75	2.84	3.82	6.82	5.65
Total	1296	1296.00	1296.00	1296.00	1296.00	1296.00
χ^2		0.31	0.53	0.70	4.19	1.97
Parameter estimates $\hat{\phi}$		0.6578	0.6598	0.6625		
$\hat{\alpha}$		0.6239	0.6221	0.6192		

Note. IP: Inflated Poisson, NB: Negative binomial, ML: Maximum Likelihood

Table 2.4 Observed and expected frequencies on the basis IP, NB and NTA distribution. [Data MaGuire et al (1957), Distribution 9].

Count per plant	Observed frequency	Expected frequency				
		IP(MM)	IP(FM)	IP(ML)	NB	NTA
0	188	188.57	187.18	188.00	185.19	187.79
1	83	81.22	82.99	81.89	89.23	85.29
2	36	38.58	38.52	38.56	32.99	34.54
3	14	12.25	11.92	12.11	10.97	11.62
4	2	2.92	2.77	2.85	3.45	3.48
5	1	0.46	0.62	0.64	1.52	1.28
Total	324	324.00	324.00	324.00	324.00	324.00
χ^2		0.33	0.36	0.55	2.36	1 26
Parameter estimates $\hat{\phi}$		0.9524	0.9283	0.9424		
$\hat{\alpha}$		0.6806	0.6982	0.6876		

Table 2.5 Observed and expected frequencies on the basis of IG, NB and Hermite distribution. [Data Greenwood and Yule (1920)]

Number of accidents	Observed frequency	Expected frequency			
		IG (ML)	IG(MM)	NB	Hermite
0	447	446.99	442.94	445.39	440.04
1	132	132.89	135.86	134.90	135.54
2	42	43.58	46.39	44.00	55.66
3	21	15.95	15.57	14.69	12.82
4	3	5.02	5.22	4.96	3.20
5	2	2.57	1.02	2.56	0.74
Total	647	647.00	647.00	647.00	647.00
χ^2		2.54	2.7	3.7	9.015
Parameter estimates $\hat{p}$		0.664	0.677		
$\hat{\alpha}$		0.9212	0.9792		

Note. IG: Inflated Geometric

Table 2.6 Observed and expected frequencies on the basis of IG and NB distribution. [Data McGuire et al. (1957), Distribution 8]

Count per plant	Observed frequency	Expected frequency		
		IG (ML)	IG(MM)	NB
0	1117	1117.00	1114.77	1114.98
1	149	151.13	155.39	154.51
2	27	23.52	24.11	22.51
3	3	4.35	1.73	3.99
Total	1296	1296.00	1296.00	1296.00
χ^2		0.96	1.54	1.35
Parameter estimates $\hat{p}$		0.84439	0.8548	
$\hat{\alpha}$		0.8873	0.9633	

CHAPTER 3

- POISSON-LINDLEY AND SOME OF ITS MIXTURE DISTRIBUTIONS

Chapter 3

Poisson-Lindley and some of its Mixture Distributions

3.1 Introduction

Poisson-Lindley distribution is a generalized Poisson distribution originally due to Lindley (1958) with pmf

$$P_x(\phi) = \frac{\phi^2(\phi + 2 + x)}{(\phi + 1)^{x+3}} \qquad x=0,1,2,\ldots \qquad (3.1.1)$$

Sankaran (1970) further investigated this distribution with application to mistakes in copying groups of random digits [data from Kemp and Kemp (1965)] and accidents to 647 women on high explosive shells in 5 week [data from Greenwood and Yule (1920)]. In both the above examples, single parameter Poisson- Lindley distribution gives a better fit than Poisson distribution. It is a special case of Bhattacharya's (1966) more complicated mixed Poisson distribution. Some mixtures of Poisson-Lindley distribution by using Gurland's generalization (1957) were studied by Borah and Deka Nath (2001) where the properties of Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distributions were investigated.

In this chapter, Poisson-Lindley distribution is further investigated. Recurrence relations for the probabilities and the factorial moments are studied. Two mixture distributions of Poisson-Lindley distribution i.e. Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distributions are investigated and their recurrence relations for factorial moments and probabilities are discussed. The parameters of the distributions are estimated by different methods of estimation. The aim of this chapter is to derive some basic properties of these three distributions and to compare them with other distributions on the basis of their fits to empirical data.

3.2 Poisson-Lindley Distribution

(a) Expression for Probabilities

The probability generating function (pgf) of Poisson-Lindley distribution may be written as

$$H(t) = \frac{\phi^2(\phi+2-t)}{(\phi+1)(\phi+1-t)^2} \tag{3.2.1}$$

Differentiating equation (3.2.1) with respect to 't', we get

$$H'(t) = \frac{\phi^2(\phi+3-t)}{(\phi+1)(\phi+1-t)^3}$$

$$\Rightarrow \qquad (\phi+1-t)^3 H'(t) = (\phi+1-t)^2 H(t) + \frac{\phi^2}{(\phi+1)} \tag{3.2.2}$$

Equating the r^{th} term in the equation (3.2.2), the following recurrence relation for probabilities may be obtained.

$$P_r = \frac{(\phi + 2 + r)}{(\phi + 1)(\phi + 1 + r)} P_{r-1} \qquad (3.2.3)$$

where $$P_0 = \frac{\phi^2(\phi + 2 - t)}{(\phi + 1)^3}$$

Putting r=1,2,3...in equation (3.2.3), the higher order probabilities may be computed easily.

(b) Factorial Moments

The factorial moment generating function (fmgf) of Poisson-Lindley distribution may be written as

$$H(t+1) = \frac{1 - \frac{t}{(\phi + 1)}}{\left(1 - \frac{t}{\phi}\right)^2} \qquad (3.2.4)$$

Differentiating the equation (3.2.4) with respect to't' and equating the coefficients of $\frac{t^r}{r!}$, the following relation for factorial moments may be obtained.

$$\mu'_{(r)} = \frac{r(\phi + r + 1)}{\phi(\phi + r)} \mu'_{(r-1)}, \qquad \text{for } r = 2,3,4,\ldots \qquad (3.2.5)$$

where $$\mu'_{(1)} = \frac{(\phi + 2)}{\phi(\phi + 1)}$$ and

$$\mu'_{(2)} = \frac{2!(\phi + 3)}{\phi^2(\phi + 1)}$$

where $\mu'_{(r)}$ stands for the rth factorial moments.

Hence variance = $\mu_2 = \frac{\phi^3 + 4\phi^2 + 6\phi + 2}{\phi^2(\phi + 1)^2}$

(c) Estimation of Parameter

The single parameter ϕ of the Poisson-Lindley distribution can be estimated by the following methods.

(i) Method of moments

Using the sample mean $\bar{x}$, the parameter ϕ of Poisson-Lindley distribution may be estimated as

$$\hat{\phi} = \frac{\left[-(\bar{x}-1)+\sqrt{(\bar{x}-1)^2+8\bar{x}}\right]}{2\bar{x}} \tag{3.2.6}$$

where $\bar{x} = \dfrac{(\phi+2)}{\phi(\phi+1)}$ [see Sankaran (1970)].

(ii) Ratio of first two frequencies

For Poisson-Lindley distribution, ϕ may also be estimated by taking ratio of first two frequencies We have

$$\frac{f_0}{N} = \frac{\phi(\phi+2)}{(\phi+1)^3} \quad \text{and} \quad \frac{f_1}{N} = \frac{\phi^2(\phi+3)}{(\phi+1)^4}$$

Eliminating ϕ between first two frequencies, we may obtain

$$\hat{\phi} = \frac{-(3f_1-f_0)+\sqrt{(3f_1-f_0)^2-4f_1(2f_1-3f_0)}}{2f_1} \tag{3.2.7}$$

3.3 Poisson-Poisson-Lindley Distribution

Poisson-Poisson-Lindley distribution may be derived by generalizing Poisson distribution [see Gurland (1957)], by using Poisson-Lindley distribution.

(a) Expression for Probabilities

By using Gurland's theorem [Gurland (1957)], the pgf of Poisson-Poisson- Lindley distribution may be written as

$$H(t) = e^{\lambda\left\{\frac{\phi^2(\phi+2-t)}{(\phi+1)(\phi+1-t)^2}-1\right\}} \tag{3.3.1}$$

Differentiating both sides of equation (3.3.1) with respect to 't' , we get

$$H'(t) = \frac{\lambda\phi^2(\phi+3-t)}{(\phi+1)(\phi+1-t)^3} H(t) \tag{3.3.2}$$

Equating the coefficient of t^r on both sides of the equation (3.3.2), we get

$$\Rightarrow \quad (r+1)(\phi+1)^3 P_{r+1} - 3(\phi+1)^2 rP_r + 3(\phi+1)(r-1)P_{r-1} - (r-2)P_{r-2} =$$

$$\lambda\phi^2\{(\phi+3)P_r - P_{r-1}\}/(\phi+1)$$

Hence the recurrence relation for probabilities may be written as

$$P_{(r+1)} = \frac{1}{(r+1)}\left[\left\{3\alpha r + \lambda\phi^2\frac{\alpha(\phi+3)}{(\phi+1)^3}\right\}P_r - \left\{3\alpha^2(r-1) + \frac{\lambda\phi^2\alpha}{(\phi+1)^3}\right\}P_{r-1} + \right.$$

$$\left. \alpha^3(r-2)P_{r-2} \ \right], \quad r=2,3,\ldots, \tag{3.3.3}$$

where $\lambda>0$, $\phi>0$, $\alpha = \frac{1}{\phi+1}$, $P_0 = e^{\lambda\left\{\frac{\phi^2(\phi+2)}{(\phi+1)^3}-1\right\}}$,

$$P_1 = \left[\frac{\lambda\phi^2(\phi+3)}{(\phi+1)^3}\right]P_0 \text{ and}$$

$$P_2 = \frac{1}{2}\left[\left\{\frac{\lambda\phi^2(\phi+3)\alpha}{(\phi+1)^3} + \frac{3r}{(\phi+1)}\right\}P_1 - \frac{\lambda\phi^2\alpha}{(\phi+1)^2}P_0\right]$$

(b) Factorial Moments

The fmgf of Poisson-Poisson-Lindley distribution may be written as

$$H(1+t)=e^{\lambda\left\{\frac{\left(1-\frac{t}{(\phi+1)}\right)}{\left(1-\frac{t}{\phi}\right)^2}-1\right\}} \qquad (3.3.4)$$

Differentiating both sides of equation (3.3.4) with respect to't' and equating the coefficient of $\frac{t^r}{r!}$, the following relation for factorial moments may be obtained.

$$\mu'_{(r+1)}=\left\{\frac{3r}{\phi}+\frac{\lambda(\varphi+2)}{\phi(\phi+1)}\right\}\mu'_{(r)}-\left\{\frac{3r(r-1)}{\phi^2}+\frac{\lambda r}{\phi(\phi+1)}\right\}\mu'_{(r-1)}+\frac{r(r-1)(r-2)}{\phi^3}\mu'_{(r-2)}$$

for r=2,3,.... (3.3.5)

where $\mu'_{(r)}$ denotes the r[th] factorial moment of Poisson-Poisson-Lindley distribution and

$$\mu'_{(1)}=\frac{\lambda(\phi+2)}{\phi(\phi+1)},$$

$$\mu'_{(2)}=\frac{\lambda^2(\phi+2)^2}{\phi^2(\phi+1)^2}+\frac{2\lambda(\phi+3)}{\phi^2(\phi+1)},$$

$$\mu'_{(3)}=\frac{\lambda^3(\phi+2)^3}{\phi^3(\phi+1)^3}+\frac{2\lambda^2(\phi+2)(5\phi+9)}{\phi^3(\phi+1)^2}+\frac{6\lambda(3\phi+8)}{\phi^3(\phi+1)},$$

$$\mu'_{(4)}=\frac{\lambda^4(\phi+2)^4}{\phi^4(\phi+1)^4}+\frac{\lambda^3(\phi+2)^2(7\phi+27)}{\phi^4(\phi+1)^3}+\frac{12\lambda^2(7\phi^2+28\phi+29)}{\phi^4(\phi+1)^2}+\frac{12\lambda(11\phi+28)}{\phi^4(\phi+1)}$$

Hence Variance $=\mu_2=\dfrac{\lambda(\phi^2+4\phi+6)}{\phi^2(\phi+1)}$

(c) Estimation of Parameters

The two parameters 'λ' and 'ϕ' of Poisson-Poisson-Lindley distribution may be estimated by the following methods.

(i) Method of moments

Using the sample mean and variance the two parameters λ and ϕ of Poisson-Poisson-Lindley distribution may be estimated as follows.

$$\hat{\phi} = \frac{-\left(2\bar{x} - S^2\right) + \sqrt{\left(2\bar{x} - S^2\right)^2 - 6\bar{x}\left(\bar{x} - S^2\right)}}{\left(\bar{x} - S^2\right)} \tag{3.3.6}$$

$$\hat{\lambda} = \frac{\hat{\phi}\left(\hat{\phi} + 1\right)\bar{x}}{\left(\hat{\phi} + 2\right)} \tag{3.3.7}$$

where $\bar{x} = \dfrac{\lambda(\phi + 2)}{\phi(\phi + 1)}$ and

$$S^2 = \frac{\lambda\left(\phi^2 + 4\phi + 6\right)}{\phi^2(\phi + 1)}.$$

(ii) Ratio of first two frequencies and the mean

By equating first two probabilities with f_0 / N and f_1 / N , we get

$$\lambda^* = \frac{f_1\left(\phi^* + 1\right)^3}{f_0 \phi^{*2} \{2\alpha\left(\phi^* + 2\right) - 1\}} \tag{3.3.8}$$

$$\phi^* = \frac{-\left(2\bar{x} - S^2\right) + \sqrt{\left(2\bar{x} - S^2\right)^2 - 6\bar{x}\left(\bar{x} - S^2\right)}}{\left(\bar{x} - S^2\right)} \tag{3.3.9}$$

where $\dfrac{f_1}{N} = \left[\dfrac{\lambda\phi^2(\phi + 3)}{(\phi + 1)^3}\right]\dfrac{f_0}{N}$

3.4 Poisson-Lindley-Poisson Distribution

Poisson-Lindley-Poisson distribution may be derived by generalizing Poisson-Lindley distribution [see Gurland (1957)] using Poisson distribution as the generalizing distribution.

(a) Expression for Probabilities

The pgf of Poisson-Lindley-Poisson distribution may be written as

$$H(t) = \frac{\phi^2\left\{\phi + 2 - e^{\lambda(t-1)}\right\}}{(\phi+1)\left(\phi+1-e^{\lambda(t-1)}\right)^2}$$

$$= A\frac{\phi + 2 - e^{\lambda(t-1)}}{\{1-\alpha e^{\lambda(t-1)}\}^2} \qquad (3.4.1)$$

where $A = \dfrac{\phi^2}{(\phi+1)^3}$ and

$$\alpha = \frac{1}{(\phi+1)}$$

Differentiating both sides of equation (3.4.1) with respect to 't', we get

$$\left(1-\alpha e^{\lambda(t-1)}\right)^3 H'(t) = Ae^{\lambda(t-1)}\lambda\left\{2\alpha(\phi+2) - 1 - \alpha e^{\lambda(t-1)}\right\}$$

$$\Rightarrow \left(1 - 3\alpha e^{\lambda(t-1)} + 3\alpha^2 e^{2\lambda(t-1)} - \alpha^3 e^{3\lambda(t-1)}\right)H'(t) = A\left\{\lambda e^{\lambda(t-1)}\frac{(\phi+3)}{(\phi+1)} - \lambda\alpha e^{2\lambda(t-1)}\right\}$$

$$(3.4.2)$$

Equating the coefficients of t^r on both sides of the equation (3.4.2), the recurrence relation for probabilities of Poisson-Lindley-Poisson distribution may be written as

$$P_{r+1}=\frac{B}{(r+1)}\left[A\left\{\frac{(\phi+3)}{(\phi+1)}\lambda e^{-\lambda}-2^{r}\alpha\lambda e^{-2\lambda}\right\}\frac{\lambda^{r}}{r!}+3\alpha e^{-\lambda}\sum_{j=1}^{r}(1-2^{j}\alpha e^{-\lambda}+3^{j-1}\alpha^{2}e^{-2\lambda})\right.$$

$$\left.(r-j+1)\frac{\lambda^{j}}{j!}p_{r-j+1}\right] \tag{3.4.2}$$

where $B=\frac{1}{\left(1-3\alpha e^{-\lambda}+3\alpha^{2}e^{-2\lambda}-\alpha^{3}e^{-3\lambda}\right)}$, $P_0=\frac{A\left(\phi+2-e^{-\lambda}\right)}{\left(1-\alpha e^{-\lambda}\right)^{2}}$ and

$$P_1=AB\lambda e^{-\lambda}\left\{\frac{(\phi+3)}{(\phi+1)}-\alpha e^{-\lambda}\right\}$$

(b) Factorial Moments

The fmgf of Poisson-Lindley- Poisson distribution may be written as

$$H(t+1)=\frac{A\left(\phi+2-e^{-\lambda t}\right)}{\left(1-\alpha e^{-\lambda t}\right)^{2}} \tag{3.4.3}$$

Expanding and equating the term $\frac{t^r}{r!}$ on both sides of equation (3.4.3), the following factorial moment recurrence relation of Poisson-Lindley-Poisson distribution may be obtained.

$$\mu'_{(r+1)}=\frac{A}{(1-\alpha)^{3}}\left\{\frac{(\phi+3)}{(\phi+1)}\lambda^{r+1}-2^{r}\lambda^{r+1}\alpha+\sum_{j=1}^{r}\left(3\alpha-3\alpha^{2}2^{j}+\alpha^{3}3^{j}\right)\binom{r}{j}\lambda^{j}\mu'_{(r-j+1)}\right\}$$

$$\text{for } r=1,2,3,\ldots \tag{3.4.4}$$

Putting r= 1,2,3 in equation (3.4.4), the higher order moments may be obtained as

$$\mu'_{(2)}=\frac{\lambda^{2}\left(\phi^{2}+4\phi+6\right)}{\phi^{2}(\phi+1)},$$

$$\mu'_{(3)} = \frac{\lambda^3(\phi^3 + 8\phi^2 + 12\phi + 48)}{\phi^3(\phi + 1)} \quad \text{and}$$

$$\mu'_{(4)} = \frac{\lambda^4(\phi^4 + 16\phi^3 + 78\phi^2 + 60\phi + 336)}{\phi^4(\phi + 1)}$$

respectively where $\mu'_{(r)}$ denotes the rth factorial moment of Poisson-Lindley-Poisson distribution. Hence the mean and the variance for Poisson-Lindley-Poisson distribution may be written as

$$\text{mean} = \mu'_1 = \frac{\lambda(\phi + 2)}{\phi(\phi + 1)} \quad \text{and}$$

$$\text{variance} = \mu_2 = \frac{\lambda^2(\phi^3 + 4\phi^2 + 6\phi + 2)}{\phi^2(\phi + 1)^2} + \frac{\lambda(\phi + 2)}{\phi(\phi + 1)}$$

(c) Estimation of Parameters

The two parameters 'λ' and 'ϕ' of Poisson-Lindley-Poisson distribution may be estimated by the method of moments. By eliminating λ between mean and variance, we obtain

$$\frac{\phi^3 + 4\phi^2 + 6\phi + 2}{(\phi + 2)^2} = \frac{\mu_2 - \mu'_1}{\mu_1'^2} \tag{3.4.5}$$

which gives an estimate for ϕ by a numerical solution by using Newton-Raphson method. On getting the estimate $\hat{\phi}$ of ϕ from equation (3.4.5), the estimate of λ may be derived as

$$\hat{\lambda} = \frac{\bar{x}\hat{\phi}(\hat{\phi} + 1)}{(\hat{\phi} + 2)}, \tag{3.4.6}$$

where $\bar{x}$ denotes the mean of the distribution.

3.5 Goodness of Fit

It is believed that the Poisson-Lindley and the two of its mixture distributions i.e Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distribution should give a reasonably good fit to some numerical data for which various modified forms of Poisson distributions was fitted earlier. Therefore, we have tried the fitting of these distributions to some published data and have compared them with other distributions as measured by χ^2 criterion. In getting the χ^2 criterion for goodness of fit, tail frequencies are grouped to obtain 5 or slightly greater than 5 for the expected frequency in each group.

In Table 3.1, we have considered the problem of accidents to 647 women on high explosive shells in 5 week period [data from Greenwood and Yule (1920)] for which Poisson distribution was fitted earlier and Poisson-Lindley was fitted by Sankaran (1970). The problem of mistakes in copying groups of random digits [data from Kemp and Kemp (1965)] is considered in Table 3.2 for which single parameter Poisson distribution was fitted. Observing the values of χ^2 and a comparison of the observed frequencies with the expected frequencies for corresponding Poisson-Lindley, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distribution in both the Table 3.1 & 3.2, it is clearly seen that the Poisson-Lindley and two of its mixture distributions describe the data very well. In Table 3.3, we have considered the number of European red mites on apple leaves [Data Bliss, (1953)] for which negative binomial distribution was fitted.

Again in Table 3.4, the data on Pyrausta nibilais is considered for which Neyman Type A distribution was fitted by Beall and Rescia (1953).

In the last table of this chapter, we considered the problem of home injuries of 122 experienced men during 11 years for which generalized Poisson distribution was fitted by Consul (1989). Because of complexity of maximum likelihood method of estimation, method of moments and ratio of first two frequencies and mean are used to estimate the parameters of these distributions.

For all the five sets of data, the relative efficiency of Poisson-Lindley, Poisson-Poisson-Lindley distribution and Poisson-Lindley-Poisson distribution are shown in Table 3.1, 3.2, 3.3, 3.4 and 3.5 respectively, by using different method of estimation..

From the following tables it is clear that there are some improvement, however small it may be, in fitting these mixture distributions i.e. Poisson-Lindley, Poisson-Poisson-Lindley distribution and Poisson-Lindley-Poisson distribution over the other distributions considered earlier. The distributions as indicated here may be used with in other situations also.

Table 3.1 Comparison of observed frequencies for accidents to 647 women on high explosive shells in 5 weeks with fitted Poisson-Lindley (PL), Poisson-Poisson-Lindley (PPL) and Poisson-Lindley-Poisson (PLP) distributions. [Data Greenwood and Yule (1920)]

No. of accident	Observed frequency	Expected Frequency					
		Poisson	PL(MM)	PL(RF)	PPL(MM)	PPL(RF)	PLP(MM)
0	447	406	441.28	452.90	442.05	450.95	444.58
1	132	189	139.83	135.15	137.79	133.15	134.78
2	42	45	45.02	38.82	46.57	43.86	46.17
3	21	7	14.88	13.86	14.57	13.46	14.85
4	3	1	4.20	4.08	4.32	3.24	4.62
≥ 5	2	1	1.79	2.19	1.70	2 34	2.00
Total	647	649	647.00	647.00	647.00	647.00	647.00
χ^2		50.57	3.84	4.39	4.04	4.41	3.56
$\hat{\theta}$			2.726	3.056	5.163	5.163	0.567
$\hat{\lambda}$					2.066	1.957	0.161

Note: PL(MM): Poisson-Lindley distribution (Method of moments)

PL(RF): Poisson-Lindley distribution (Ratio of first two frequencies)

PPL(MM): Poisson-Poisson-Lindley distribution (Method of moments)

PPL(RF): Poisson-Poisson-Lindley distribution (Ratio of first two frequencies)

PLP(MM): Poisson-Lindley-Poisson distribution (Method of moments)

Table 3.2 Distribution of Mistakes in copying groups of random digits with expected frequencies obtained by Poisson, Poisson-Lindley (PL), Poisson-Poisson-Lindley (PPL) and Poisson-Lindley-Poisson (PLP) distribution. [Data Kemp and Kemp (1965)]

No. of errors per group	Observed frequency	Expected Frequency					
		Poisson	PL(MM)	PL(RF)	PPL(MM)	PPL(RF)	PLP(MM)
0	35	27.4	33.05	41.34	32.83	39.87	35.53
1	11	21.5	15.27	12.99	15.22	12.53	15.69
2	8	8.4	6.74	3.46	7.06	4.87	7.02
3	4	2.2	2.89	1.82	2.99	1.78	2.91
4	2	.4	1.21	0.39	1.19	0.95	1.15
Total	60	59.9	59.16	60.00	59.29	60.00	59.29
χ^2		15.13	2.48	16.49	2.33	6.72	2.58
$\hat{\theta}$			1.7421	2.839	3.930	3.930	0.244
$\hat{\lambda}$					2.559	1.735	0.106

Table 3.3 Comparison of observed frequencies for counts of the number of European red mites on apple leaves with expected negative binomial, Poisson-Lindley (PL), Poisson-Poisson-Lindley (PPL) and Poisson-Lindley-Poisson (PLP) distributions. [Data Bliss (1963)].

No.of mites per leaf	Observed frequency	Expected Frequency					
		N.B.	PL(MM)	PL(RF)	PPL(MM)	PPL(RF)	PLP(MM)
0	70	67.49	67.189	71.832	69.879	66.146	68.675
1	38	39.03	38.884	38.995	35.391	35.908	36.944
2	17	20.86	21.262	20.068	21.016	21.977	21.213
3	10	10.97	11.206	9.972	11.626	12.469	11.405
4	9	5.66	5.755	4.834	6.116	6.707	5.911
5	3	2.90	2.899	2.300	3.096	3.465	2.986
6	2	1.48	1.439	1.079	1.521	1.734	1.480
7	1	0.75	0.706	0.500	0.728	0.845	0.723
8	0	0.76	0.660	0.420	0.627	0.749	0.663
Total	150	150.00	150.00	150.00	150.00	150.00	150.00
	χ^2	1.9275	1.6853	2.5795	1.4898	1.9906	1.4909
	$\hat{\theta}$		1.258	1.386	2.472	2.472	0.357
	$\hat{\lambda}$				2.201	2.359	0.236

Table 3.4 Observed and fitted Poisson-Lindley (PL), Poisson-Poisson-Lindley (PPL) and Poisson-Lindyey-Poisson (PLP) distribution. [Data on the Pyrausta nibilais, to which NTA was fitted by Beall & Rescia (1953)].

No.of insects	Observed frequency	Expected Frequency					
		NTA	PL(MM)	PL(FM)	PPL(MM)	PPL(FM)	PLP(MM)
0	33	37.8	31.485	35.968	32.503	33.881	32.212
1	12	5.6	14.156	13.079	12.796	12.320	13.191
2	6	5.2	6.090	4.592	6.079	5.668	6.094
3	3	3.5	2.542	1.572	2.703	2.453	2.642
4	1	1.9	1.038	0.529	1.147	1.016	1.107
5	1	2.0	0.689	0.260	0.772	0.662	0.754
Total	56.	56.00	56.00	56.00	56.00	56.00	56.00
	χ^2	9.04	0.437	0.791	0.065	0.178	0.037
	$\hat{\theta}$		1.808	2.378	3.176	3.176	0.487
	$\hat{\lambda}$				1.922	1.775	0.218

Table 3.5 Comparison of observed frequencies for home injuries of 122 experience men during 11years (1937-1947) with expected Poisson-Lindley (PL), Poisson-Poisson-Lindley (PPL) and Poisson-Lindyey-Poisson (PLP) and GPD [Consul (1989)] frequencies.

No.of	Observed	Expected Frequency					
injuries	frequency	GPD	PL(MM)	PL(RF)	PPL(MM)	PPL(RF)	PLP(MM)
0	58	57.22	59.753	53.869	57.051	57.083	56.430
1	34	34.41	31.693	31.579	33.443	33.437	34.530
2	14	16.64	15.956	17.479	17.486	17.476	17.351
3	8	7.59	7.761	9.321	8.158	8.150	7.911
4	6	6.14	3.683	4.843	3.525	3.521	3.411
5	2		3.154	4.909	2.337	2.333	2.367
Total	122	122	122	122	122	122	122
	χ^2	1.09	0.664	0.315	1.503	1.505	1.554
	$\hat{\theta}$		1.434	1.233	3.976	3.976	0.094
	$\hat{\lambda}$				3.259	3.256	0.048

CHAPTER 4

- A STUDY ON INFLATED POISSON-LINDLEY DISTRIBUTION

Chapter 4

A study on Inflated Poisson-Lindley Distribution

4.1 Introduction

To serve the probabilistic description of an experiment with a slight inflation of probability at point zero, the inflated Poisson-Lindley distribution is studied in this chapter. The recurrence relations for moments and probabilities of the inflated Poisson-Lindley (IPL) distribution are derived. Attempt has been made to obtain the recurrence relations without derivatives, since these forms are easy to handle on computer. Borah and Deka Nath (2001) studied and fitted the IPL distribution to some well known data for empirical comparison.

4.2 Recurrence Relation for Probabilities

The probability generating function (pgf) of the IPL distribution may be written as

$$H(t) = \beta + \alpha g(t) \qquad (4.2.1)$$

where $g(t)=\dfrac{\phi^2(\phi+2-t)}{(\phi+1)(\phi+1-t)^2}$ is the pgf of Poisson-Lindley (PL) distribution and $\alpha+\beta=1,\ 0<\beta<1$ and $\phi>0$.

It is also possible to take the parameter β less than zero, provided $\beta + \alpha p_0 \geq 0$, where $p_0 = P(X = 0)$.

Differentiating equation (4.2.1) with respect to 't' and equating the coefficients of t^r from both sides of the equation, we have the recurrence relation for probabilities

$$P(X = r) = \frac{(\phi + 2 + r)}{(\phi + 1)(\phi + 1 + r)} P(X = r - 1) \ , \qquad r=2,3,\ldots \qquad (4.2.2)$$

where $P(X = 0) = \beta + \dfrac{\alpha\phi^2(\phi + 2)}{(\phi + 1)^3}$, $\qquad P(X = 1) = \dfrac{\alpha\phi^2(\phi + 3)}{(\phi + 1)^4}$

After some suitable transformation of equation (4.2.2), we may have

$$P(X = r) = \frac{\alpha\phi^2(\phi + 2 + r)}{(\phi + 1)^{3+r}}, \qquad r=1,2,3,\ldots \qquad (4.2.3)$$

where $P(X = 0) = \beta + \dfrac{\alpha\phi^2(\phi + 2)}{(\phi + 1)^3}$.

4.3 Graphical Representation of IPL Distribution

To study the behaviour of the IPL distribution with varying values of ϕ and $\beta = 1 - \alpha$, the probabilities for possible values of X are computed by using above equations (4.2.2) or (4.2.3) and accordingly different graphs may be drawn for various values of the two parameters.

Fig. 4.1 contains four graphical representation for β= -0.1 and ϕ=1,2,5,7. Similarly, Fig. 4.2 contains four graphs for β= -0.4 and ϕ =2,3,5,9 respectively. It clearly indicates from the graphs that for the changes in the values of ϕ there are significant differences in the probability distribution i.e. as ϕ

increases, the value of P(X=0) increases and the probabilities for all other values of X decrease.

For further study of the effect of the changes in the values of β on the behaviour of IPL distribution, Figures 4.3 and 4.4 are shown. Fig. 4.3 contains two graphs for β=0.1 and ϕ =1 & 3 and Fig 4.4 contains two graphs for β=0.9 and ϕ =1 & 5 respectively. It is clear from these graphs of Fig. 4.3 and Fig. 4.4 that when β increases, the value of P(X=0) increases and the probabilities for all other values of X decrease and the model always remains L-shaped. Thus the tail becomes more and more heavier and longer with the decrease in the value of ϕ. It is also clear from Figure 4.4 that the probabilities on the right hand side tail sharply decline when β closes to unity and ϕ takes any value.

Again Fig. 4.5 contains four graphs for ϕ =2 and β=-0.8,-0.6,-0.2,0 4 and Fig. 4.6 contains four graphs for ϕ =9 and β=-0.8,-0.6,0.2,0.6 respectively. From Fig. 4.5, it is seen that the model looses its L-shaped form when β and ϕ are both small. But from Fig. 4.6, it is observed that there is a similar effect like in the Fig 4.4 that, when the value of ϕ is large and the value of β is slowly increased, then P(X=0) increases and the probabilities for other values of X decrease sharply giving a L-shaped form to the model like the Geeta distribution [Consul (1990)].

Figure 4.1 Graphs of probability distribution for $\beta = -0.1$ and $\phi = 1,2,5,7$ respectively for IPL distribution.

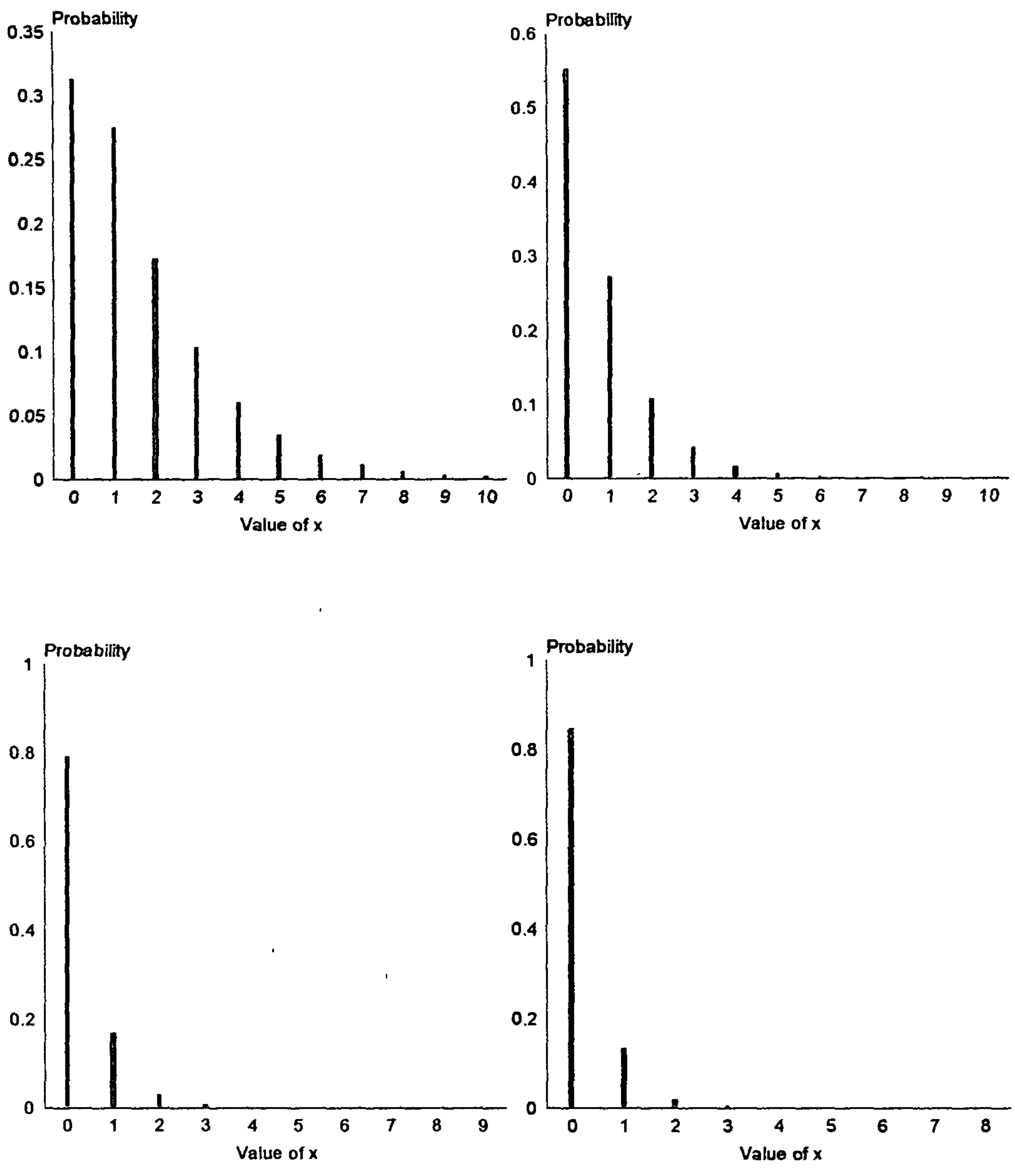


Figure 4.2 Graphs of probability distribution for $\beta = -0.4$ and $\phi = 2,3,5,9$ respectively for IPL distribution

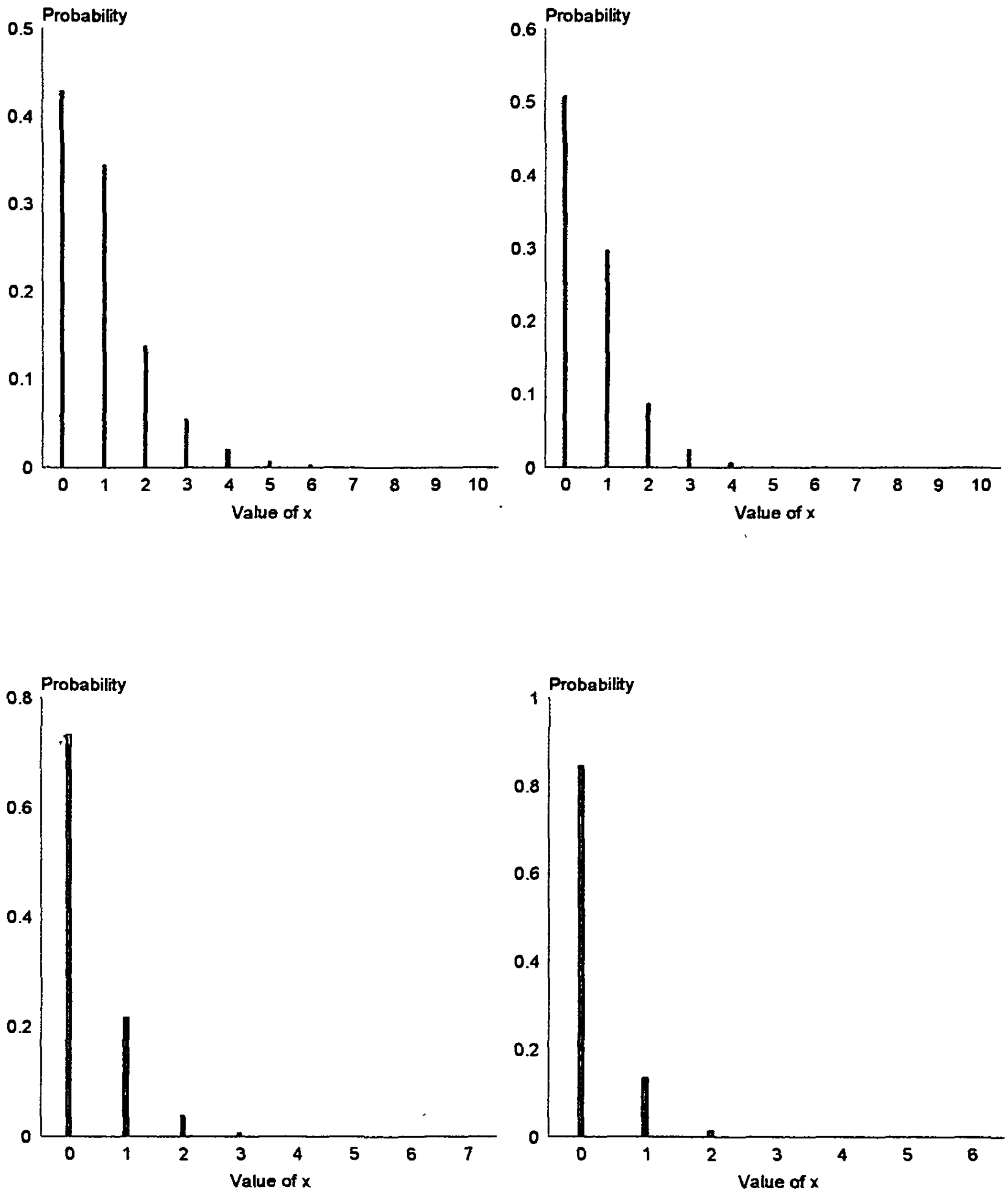
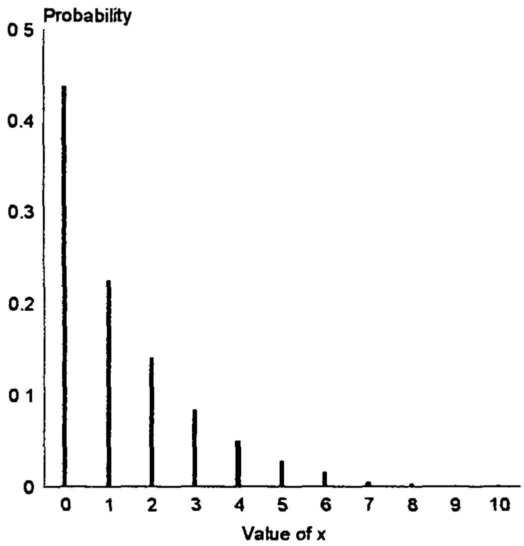


Figure 4.3 Graphs of probability distribution for $\beta = 0.1$ and $\phi = 1,3$ respectively for IPL distribution.



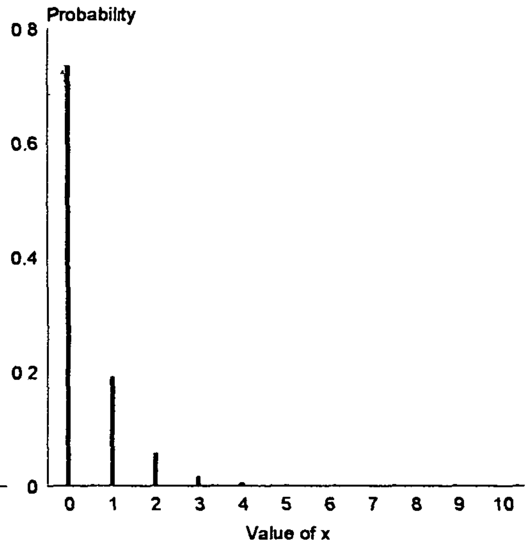
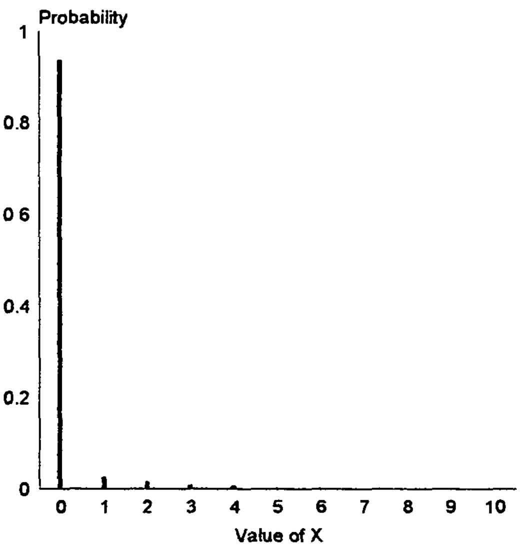


Figure 4.4 Graphs of probability distribution for $\beta = 0.9$ and $\phi = 1,\ 5$ respectively for IPL distribution.



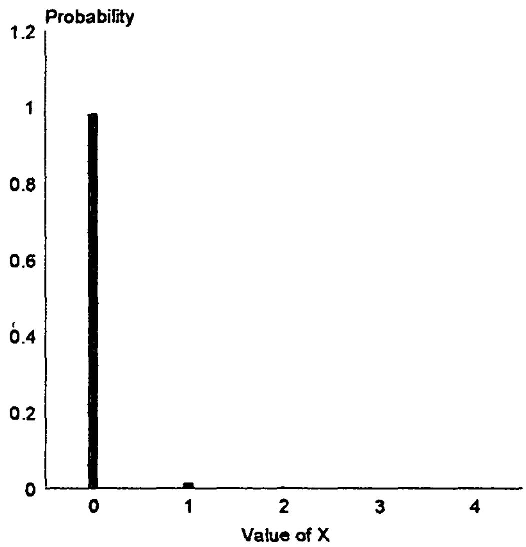


Figure 4.5 Graphs of probability distribution for $\phi = 2$ and $\beta = -0.8, -0.6, -0.2, 0.4$ respectively for IPL distribution.

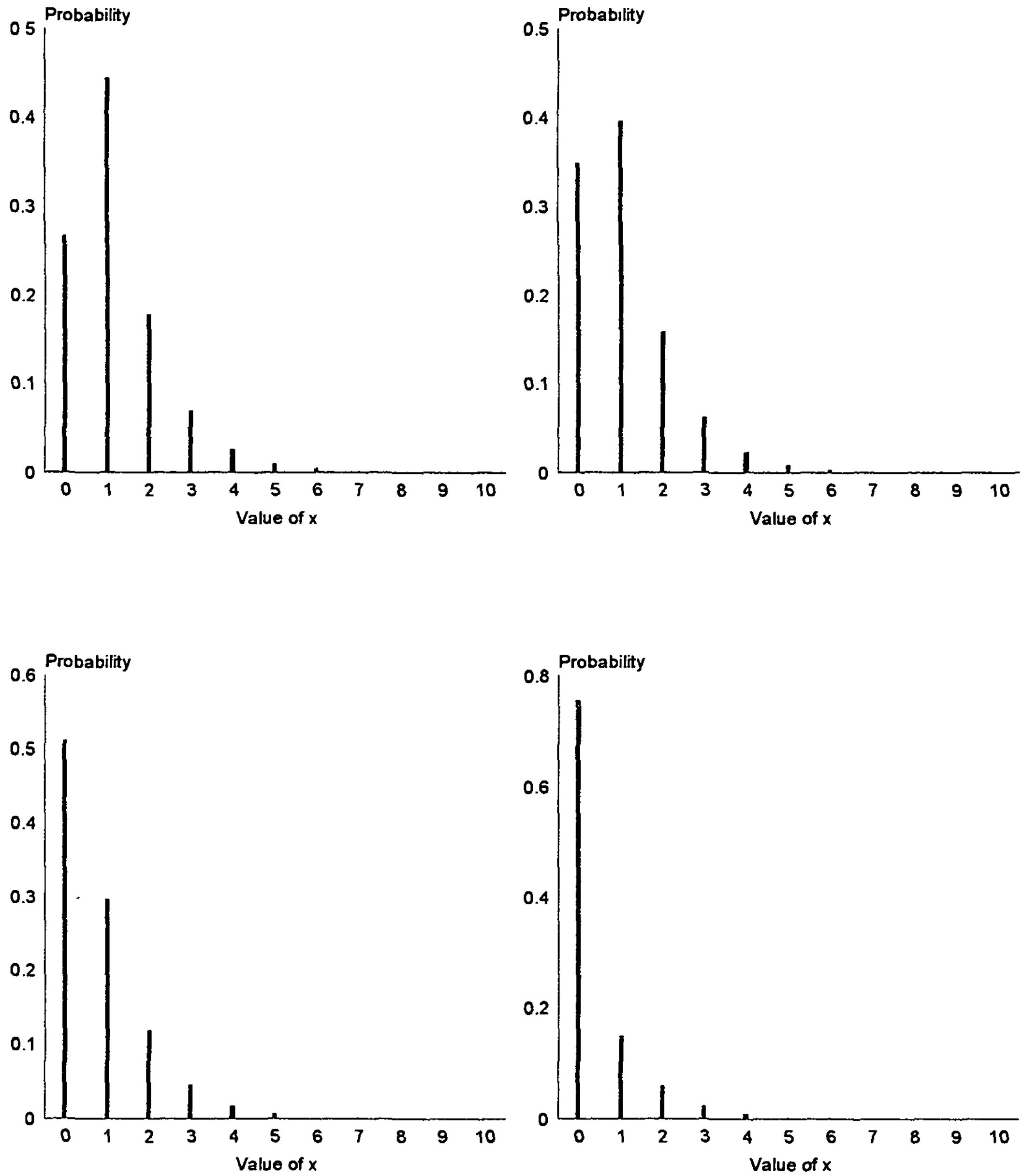
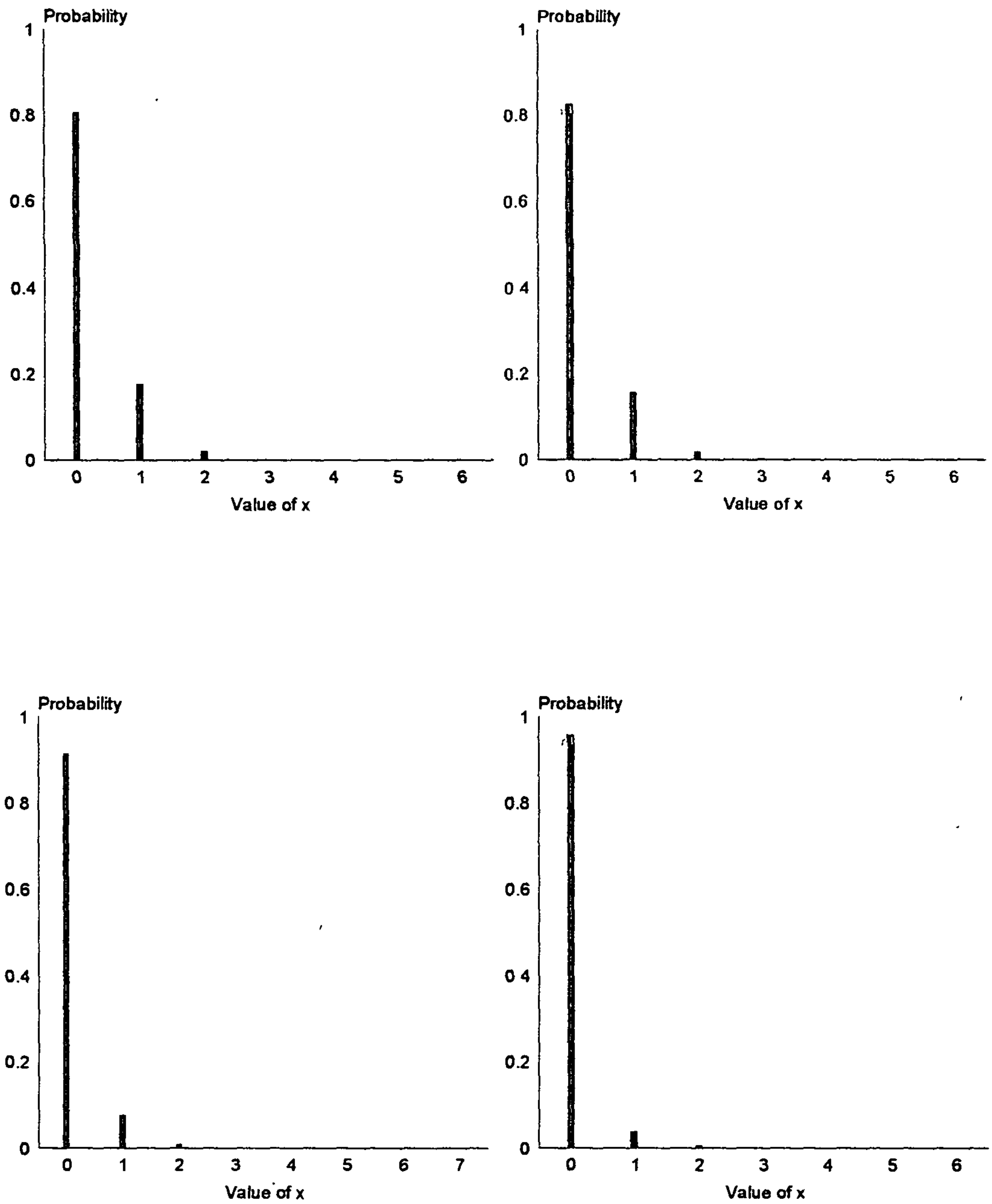


Figure 4.6 Graphs of probability distribution for $\phi = 9$ and $\beta = -0.8, -0.6, 0.2, 0.6$ respectively for IPL distribution.



4.4 Recurrence Relation for Raw Moments

The moment generating function (mgf) of IPL distribution may be written as

$$m(t) = \beta + \alpha M(t) \tag{4.4.1}$$

where $M(t) = \dfrac{\phi^2(\phi + 2 - e^t)}{(\phi+1)(\phi+1-e^t)^2}$ is the mgf of the PL distribution [see chapter 3].

Differentiating equation (4.4.1) with respect to 't' and equating the coefficient of $\dfrac{t^r}{r!}$, we get the raw moments recurrence relation for the IPL distribution as

$$\mu_r' = \frac{\alpha\{(\phi+3)-2^r\}}{\phi(\phi+1)} + \sum_{j=0}^{r-1} \frac{(3a - 3.2^{j+1}a^2 + 2^{j+1}a^3)}{(1-a)^3}\binom{r}{j+1}\mu_{r-j}', r>1 \tag{4.4.2}$$

where $a = \dfrac{1}{(\phi+1)}$

$$\mu_1' = \frac{\alpha(\phi+2)}{\phi(\phi+1)}, \tag{4.4.3}$$

and $$\mu_2' = \frac{\alpha(\phi^2+4\phi+6)}{\phi^2(\phi+1)} \tag{4.4.4}$$

where μ_r' denotes the r[th] raw moment of IPL distribution. The central moments of IPL distribution which can be obtained from the raw moments are given bellow

$$\mu_2 = \frac{\alpha\{\phi^3 + 4\phi^2 + 6\phi + 2 + \beta(\phi+2)^2\}}{\phi^2(\phi+1)^2} \tag{4.4.5}$$

$$\mu_3 = \frac{\alpha}{\phi^3(\phi+1)^3}\{\phi^5 + 7\phi^4 + 22\phi^3 + 32\phi^2 + 18\phi + 4 + \beta(3\phi^4 + 17\phi^3 + 36\phi^2 +$$

$$30\phi + 4\ ) + 2\beta^2(\phi^3 + 6\phi^2 + 12\phi + 18)\ \} \tag{4.4.6}$$

$$\mu_4 = \frac{\alpha}{\phi^4(1+\phi)^4}\{\phi^7 + 5\phi^6 + 97\phi^5 + 258\phi^4 + 406\phi^3 + 314\phi^2 + 48\phi + 24 + \beta$$

$$(4\phi^6 + 36\phi^5 + 145\phi^4 + 312\phi^3 + 400\phi^2 + 384\phi + 48) + 3\beta^2(2\phi^5$$

$$+15\phi^4 + 44\phi^3 + 52\phi^2) + 3\beta^3(\phi^4 + 8\phi^3 + 24\phi^2 + 32\phi + 16)\} \qquad (4.4.7)$$

Putting $\beta = 0$ in equation (4.4.5) the variance of PL distribution may be obtained [see Borah and Deka Nath (2001)].

4.5 Skewness and Kurtosis of IPL Distribution

The expression for the coefficient of skewness and kurtosis can be written as follows in terms of ϕ and α.

$$\gamma_1 = +\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{A + \beta B + \beta^2 C}{\sqrt{\alpha D^3}} \qquad (4.5.1)$$

where $A = \phi^5 + 7\phi^4 + 22\phi^3 + 32\phi^2 + 18\phi + 4$

$B = 3\phi^4 + 17\phi^3 + 36\phi^2 + 30\phi + 4$

$C = \phi^3 + 6\phi^2 + 12\phi + 18$

$D = \phi^3 + 4\phi^2 + 6\phi + 2 + \omega(\phi + 2)^2$

and $$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{E + \beta F + 3\beta^2 G + 3\beta^3 H}{\alpha D^2} \qquad (4.5.2)$$

where $E = \phi^7 + 2\phi^6 + 73\phi^5 + 174\phi^4 + 256\phi^3 + 152\phi^2 - 24\phi + 12$

$F = 7\phi^6 + 54\phi^5 + 181\phi^4 + 312\phi^3 + 34\phi^2 + 264\phi + 12$

$G = 4\phi^5 + 30\phi^4 + 62\phi^3 + 112\phi^2 + 32\phi$

$H = 2\phi^4 + 16\phi^3 + 52\phi^2 + 64\phi + 32$

It is clear from the above expression of γ_1 that for any given value of $\phi>0$ and β is close to unity, the skewness is infinitely large and it becomes smaller and smaller as the value of β decreases. The IPL distribution is leptokurtic as the value of γ_2 is positive for all values of $\phi>0$ and $0<\beta<1$, which is also clear from the simulated results of $\frac{\beta_2-3}{\beta_1}$:

β :	.5	.5	.5	.5	.5
ϕ :	.0001	.2	1	10	20
$\frac{\beta_2-3}{\beta_1}$:	11.432	22.06	58.49	202232	4151833

From the above simulated result, it is seen that the value of the ratio $\frac{\beta_2-3}{\beta_1}$ tends to ∞ as ϕ becomes larger. This wider limit of $\frac{\beta_2-3}{\beta_1}$ indicates greater flexibility of the IPL distribution.

4.6 Recurrence Relation for Factorial Moments

The factorial moment generating function (fmgf) of IPL distribution may be written as

$$H(1+t) = \beta + \frac{\alpha\left(1-\frac{t}{a}\right)}{\left(1-\frac{t}{\phi}\right)^2} \tag{4.6.1}$$

where $a = \frac{1}{\phi + 1}$

Differentiating equation (4.6.1) with respect to 't' and equating the coefficients of $\frac{t^r}{r!}$, the following recurrence relation for factorial moments may be obtained.

$$\mu'_{(r+1)} = \frac{3}{\phi}\mu'_{(r)} - \frac{3r(r-1)}{\phi^2}\mu'_{(r-1)} + \frac{r(r-1)(r-2)}{\phi^3}\mu'_{(r-2)} \qquad (4.6.2)$$

where $\mu'_{(1)} = \frac{1!\alpha(\phi+2)}{\phi(\phi+1)}$

$$\mu'_{(2)} = \frac{2!\alpha(\phi+3)}{\phi^2(\phi+1)}$$

$$\mu'_{(3)} = \frac{3!\alpha(\phi+4)}{\phi^3(\phi+1)}$$

where $\mu'_{(r)}$ is the rth factorial moment of IPL distribution. The recurrence relation in equation (4.6.2) of factorial moments may also be written as

$$\mu'_{(r)} = \frac{r!\alpha(\phi+r+1)}{\phi^r(\phi+1)} \qquad (4.6.3)$$

4.7 Estimation of Parameters

The parameters ϕ and β of IPL distribution can be estimated by using the following methods.

(a) Method of Maximum Likelihood (ML)

Since IPL distribution is a zero modified distribution, one of the ML equations is

$$\hat{\beta} + \hat{\alpha}\frac{\hat{\phi}^2(\hat{\phi}+2)}{(\hat{\phi}+1)^3} = \frac{n_0}{N} \tag{4.7.1}$$

where $\frac{n_0}{N}$ is the observed proportion of zeros. As it is also a power series distribution so the other ML equation will be

$$\bar{x} = \hat{\alpha}\frac{(\hat{\phi}+2)}{\hat{\phi}(\hat{\phi}+1)} \tag{4.7.2}$$

Eliminating $\hat{\beta}$ from equation (4.7.1) and (4.7.2) respectively, we may have

$$\frac{\hat{\phi}(\hat{\phi}+1)}{(\hat{\phi}+2)}\bar{x} - \frac{\hat{\phi}^2}{(\hat{\phi}+1)^2} = 1 - \frac{n_0}{N} \tag{4.7.3}$$

$\hat{\phi}$ can be estimated from equation (4.7.3) by using Newton Raphson method and then $\hat{\beta}$ may be estimated from equation (4.7.1)

(b) Methods of Moments

The parameters ϕ and β may be estimated from the first two raw moments μ_1' and μ_2' from equation (4.4.3) and (4.4.4).

Thus $$\tilde{\phi} = \frac{(2\mu_1' - \mu_2') + \sqrt{(\mu_2' - 2\mu_1' + 2\mu_1'\mu_2')}}{(\mu_2' - \mu_1')} \tag{4.7.4}$$

and $$\tilde{\beta} = 1 - \frac{\tilde{\phi}(\tilde{\phi}+1)\mu_1'}{(\tilde{\phi}+2)} \tag{4.7.5}$$

(c) Ratio of first two Frequencies and Mean

Eliminating β between first two frequencies i.e. $\frac{n_1}{N}$ and $\frac{n_2}{N}$, we get

$$\phi' = \left(\frac{n_1}{2n_2} - 2\right) + \sqrt{\left\{\left(2 - \frac{n_1}{2n_2}\right)^2 - \left(3 - \frac{4n_1}{n_2}\right)\right\}}, \tag{4.7.6}$$

where $\frac{n_1}{N} = \frac{\alpha(\phi+3)\phi^2}{(\phi+1)^4}$ and $\frac{n_2}{N} = \frac{\alpha(\phi+4)\phi^2}{(\phi+1)^5}$ and β' may be estimated from the following expression.

$$\beta' = 1 - \frac{\phi'(\phi'+1)\bar{x}}{(\phi'+2)} \tag{4.7.7}$$

4.8 Fitting of IPL Distribution to Data

As the IPL distribution has only two parameters and has a simple form so it may be applied in different fields such as biology and ecology, social information, genetic and so on which are discussed below.

Biology and Ecology

For the fitting of IPL distribution, in Table 4.1, we have considered the Student's historic data on Haemocytometer counts of yeast cells for which Gegenbauer distribution (GD) was fitted by Borah (1984) (using method of moments). In Tables 4.2 and Table 4.3, we have considered two data sets of Beall (1940), for which generalized Poisson distribution (GPD) was fitted by Jain (1975) (by using MLE). It is observed from the following Tables 4.1, 4.2 and 4.3 that ML method gives better result in all the cases and there is some improvement however small it may be, in fitting of IPL distribution over the other distributions considered earlier. In case of Table 4.3, the method of ratio of first two frequencies with mean does not give better fit, as the computed χ^2 value is quite large. Hence the result is not reported in this case.

Strikes in industries

Kendall (1961) considered the observed data on the number of strikes in 4-week period in four leading industries in U.K. during 1948-1959 and concluded that the aggregate data for the four industries agrees with Poisson law but that it did not hold well for the individual industries. The IPL distribution has been fitted to the observed data for the four individual industries and the results are given in Table 4.4 along with the expected GPD frequencies [Consul (1989)]. From the χ^2 values it is clear that the pattern of strikes in vehicle manufacture, ship building and transport industries follow IPL and GPD model and IPL distribution gives better fit than the GPD model for coal mining industries.

Genetics

Chromosome interchanges in organic cell is produced by irradiation of X-rays. Feller (1968) showed that the distribution of number of cells with exactly k interchanges should follow Poisson distribution for which he had given Catcheside et al. (1946)'s data. Consul (1989) showed that GPD model provides better fit in every case than the Poisson distribution. It is observed from Table 4.5 that IPL distribution also provides better fit, except the last case. However the χ^2 for IPL distribution in that case is lower than the significant value.

Table 4.1 Haemocytometer Counts of Yeast Cells

No. of Yeast cells per square	Observed frequency	Expected frequency		
		IPL (ML)	IPL (MM)	GD (Borah1984)
0	213	213.00	210.46	214.15
1	128	127.59	131.14	123.00
2	37	40.91	40.76	44.88
3	18	12.82	12.39	13.36
4	3	3.95	3.71	3.55
5	1	1.2	1.09	0.86
6	0	0.53	0.45	0.20
Total and Estimates	400	400.00 $\hat{\phi}$=2.669 $\hat{\beta}$=-0.431 χ^2=1.037	400.00 $\tilde{\phi}$=2.774 $\tilde{\beta}$=-0.497 χ^2=1.53	400.00 χ^2=2.8342

Table 4.2 Fit of distribution on Pyrausta niblilalis in 1937 [data of Beall (1940)]

No. of Insects	Observed Frequency	Expected frequency			
		IPL (ML)	IPL (MM)	IPL (FM)	GPD (Jain1975)
0	33	33.00	32.07	34.08	32.46
1	12	12.41	13.47	11.23	13.47
2	6	5.84	6.00	5.61	5.60
3	3	2.66	2.59	2.71	2.42
4	1	1.18	1.096	1.28	1.08
5	1	0.91	0.774	1.09	0.97
Total and Estimates	56	56.00 $\hat{\phi}$=1.588 $\hat{\beta}$=0.1406 χ^2=0.029	56.00 $\tilde{\phi}$=1.719 $\tilde{\beta}$=0.0573 χ^2=0.215	56.00 ϕ'=1.449 β'=0.228 χ^2=0.096	56.00 χ^2=0.25

Note: IPL: Inflated Poisson-Lindley distribution, ML: Maximum likelihood
MM: Method of moments, FM: Ratio of first two frequencies and mean
GPD: Generalized Poisson Distribution.

Table 4.3 Fit of distribution of Corn Borer [data of Beall (1940)]

Corn borer Per hill	Observed frequency	Expected frequency IPL (ML)	IPL (MM)	GPD (Jain 1975)
0	43	42.99	44.99	43.91
1	35	32.12	30.39	32.00
2	17	19.45	18.81	19.11
3	11	11.31	11.19	10.88
4	5	6.40	6.47	6.12
5	4	3.55	3.66	3.44
6	1	1.94	2.04	1.94
7	2	1.05	1.12	1.10
8	2	1.19	1.3	1.50
Total	120	120.00	120.00	120.00
Parameter estimates		$\hat{\phi}$=1.0587 $\hat{\beta}$=-0.5696 χ^2=0.577	$\tilde{\phi}$=1.0715 $\tilde{\beta}$=-0.0087 χ^2=0.995	χ^2=0.87

Table 4.4 Comparison of observed frequencies of the Number of outbreaks of strike in four leading industries in the U.K. during 1948-1959 with the expected IPL and GPD frequencies.

Number of outbreaks	Coal mining			Vehicle manufacture			Ship building			Transport		
	Obs.	IPL	GPD	Obs.	IPL	GPD	Obs.	IPL	GPD	Obs.	IPL	GPD
0	46	46.00	50.01	110	110.00	109.82	117	117.00	116.74	114	114.00	114.84
1	76	77.84	66.77	33	32.98	33.36	29	29.78	30.22	35	33.25	33.88
2	24	22.93	31.23	9	9.40	9.24	9	7.07	6.97	4	6.95	7.27
3	9	6.62	7.23	3	2.63	3.58	0	1.65	0.88	2	1.43	2.61
≥ 4	1	2.61	0.76	1	0.99		1	0.50		1	0.37	1.69
Total and estimates	156	156	156	156	156	156	156	156	156	156	156	156
	$\hat{\phi} = 2.963$			$\hat{\phi} = 3.084$			$\hat{\phi} = 3.828$			$\hat{\phi} = 4.43$		
	$\hat{\beta} = -1.351$			$\hat{\beta} = -0.016$			$\hat{\beta} = -0.037$			$\hat{\beta} = -0.27$		
	χ^2=0.157,		4.25	χ^2=0.056,		0.06	χ^2=1.16,		1.19	χ^2=2.14,		2.27

CHAPTER 5

- THE SHORT POISSON-POISSON-LINDLEY DISTRIBUTION

Chapter 5

The Short Poisson-Poisson-Lindley Distribution

5.1 Introduction

Cresswell & Froggatt (1963) derived a model which was a convolution of Poisson distribution and a Neyman Type A distribution. They called it 'Short', as opposite to the two-parameter 'Long' Neyman Type A distribution. The name 'Short' appears to relate the tails of the distribution. Kemp (1967) considered the properties, recurrence relations for probabilities and fitted this 'Short' distribution to accident data.

Here our distribution, namely the Short Poisson-Poisson-Lindley (SPPL) distribution is a convolution of Poisson distribution and Poisson-Poisson-Lindley distribution. The SPPL distribution which is an extension of Poisson-Poisson-Lindley distribution was studied by Deka Nath and Borah (2000).

Model derivation from accident data

While deriving the 'Short' distribution from accident data, four assumptions were made by Cresswell & Froggatt. In the same manner, the SPPL distribution has been derived by considering the following assumptions:

(i) Every driver is liable to a spell – these are periods of time during which his performance is sub-standard so that he is liable to incur accidents. The number of spells in a given time period is assumed to be Poisson variable with parameter λ_1.

(ii) All drivers are equally liable to the occurrence of a spell.

(iii) The probability of an accident occurring within a spell is constant and not dependent on the particular driver and it is assumed to have a Poisson-Lindley distribution with constant parameter θ.

(iv) Lastly, accidents can occur outside a spell and such accidents are independently distributed as Poisson distribution with parameter λ_2 over the given time period.

It is generally observed that the derivation of probability mass function (pmf) for some generalized mixture distributions seems to be complicated. So, the pgf of SPPL distribution has been obtained by using Levy's theorem [see Feller (1957)], which may be written as

$$H(t) = \exp[\{\lambda_1 (g(t) - 1) + \lambda_2 (t - 1)\}], \tag{5.1.1}$$

where $H(t)$ converges for $|t| \leq 1$, 't' is the generating parameter and $g(t) = \dfrac{\theta^2 (\theta + 2 - t)}{(\theta + 1)(\theta + 1 - t)^2}$ denotes the pgf of Poisson-Lindley distribution [see chapter 3]. Hence equation (5.1.1) may be written as

$$H(t)=\exp\left[\lambda_1\left\{\frac{\theta^2(2+\theta-t)}{(\theta+1)(\theta+1-t)^2}-1\right\}+\lambda_2(t-1)\right], \qquad (5.1.2)$$

where λ_1, $\lambda_2>0$ and $\theta>0$

In this chapter, we have studied the recurrence relation for probabilities and factorial moments of the SPPL distribution. The limiting distributions of SPPL distribution are discussed. The parameters are estimated by a composite method i.e. by using the ratio of first two frequencies and first two moments. To illustrate the various applications of this distribution, it is fitted to number of data sets. Firstly, we have considered the number of accidents sustained by a group of 708 bus drivers over a period of 3 years. Secondly, we have considered the number of accidents to 647 women on high explosive shells in 5 week periods. Thirdly, the number of accidents (home injuries) of 122 experienced men during 11 years period was considered and lastly we considered number of accidents of 122 experienced shunting men over a period of 11 years. In all the cases the SPPL distribution provides a better fit to the observed data.

5.2 Expression for Probabilities

Taking logarithm of equation (5.1.2), we have

$$\log H(t)=\lambda_1\left\{\frac{\theta^2(\theta+2-t)}{(\theta+1)(\theta+1-t)^2}-1\right\}+\lambda_2(t-1) \qquad (5.2.1)$$

Differentiating the equation (5.2.1) with respect to 't', we get

$$\frac{H'(t)}{H(t)}=\lambda_1\frac{\theta^2(\theta+3-t)}{(\theta+1)(\theta+1-t)^3}+\lambda_2$$

$$\Rightarrow \qquad (\theta+1-t)^3 H'(t) = \lambda_1 \frac{\theta^2(\theta+3-t)}{(\theta+1)} H(t) + \lambda_2(\theta+1-t)^3 H(t)$$

$$\Rightarrow \qquad \left\{(\theta+1)^3 - 3(\theta+1)^2 t + 3(\theta+1)\ t^2 - t^3\right\} H'(t) = \lambda_1 \frac{\theta^2(\theta+3-t)}{(\theta+1)} H(t) +$$

$$\lambda_2 \left\{(\theta+1)^3 - 3(\theta+1)^2 t + 3(\theta+1)t^2 - t^3\right\} H(t) \qquad (5.2.2)$$

Equating the coefficient of t^r on both sides of equation (5.2.2), we get the following recurrence relation.

$$P_{r+1} = \frac{1}{r+1}\left[\left\{\frac{\lambda_1\theta^2(\theta+3)}{(\theta+1)^4} + \frac{3r}{(\theta+1)} + \lambda_2\right\}P_r - \left\{\frac{\lambda_1\theta^2}{(\theta+1)^4} + \frac{3(r-1)}{(\theta+1)^2} + \frac{3\lambda_2}{\theta+1}\right\}P_{r-1}\right.$$

$$\left. + \left\{\frac{3\lambda_2}{(\theta+1)^2} + \frac{r-2}{(\theta+1)^3}\right\}P_{r-2} - \frac{1}{(\theta+1)^3}P_{r-3}\right], \qquad (5.2.3)$$

where $P_0 = exp\left[\lambda_1\left\{\frac{\theta^2(\theta+2)}{(\theta+1)^3} - 1\right\} - \lambda_2\right]$ (5.2.4)

$$P_1 = \left\{\lambda_1 \frac{\theta^2(\theta+3)}{(\theta+1)^4} + \lambda_2\right\} P_0 \qquad (5.2.5)$$

Putting r=1,2,3.. in equation (5.2.3), the higher order probabilities may be derived.

5.3 Expression for Factorial Moments

The factorial moment generating (fmg) function of the SPPL distribution may be written as

$$H(1+t) = \exp\left\{\lambda_1 \frac{1 - \frac{t}{(\theta+1)}}{\left(1-\frac{t}{\theta}\right)^2} + \lambda_2 t\right\} \qquad (5.3.1)$$

Taking logarithm of both sides of the equation (5.3.1) and differentiating it with respect to 't', we get the following relation

$$\frac{H'(1+t)}{H(1+t)}\left(1-\frac{t}{\theta}\right)^3 = \lambda_1 \frac{(\theta+2-t)}{\theta(\theta+1)} + \lambda_2\left(1-\frac{t}{\theta}\right)^3 \qquad (5.3.2)$$

$\Rightarrow$

$$\left(1-\frac{3t}{\theta}+\frac{3t^2}{\theta^2}-\frac{t^3}{\theta^3}\right)H'(1+t) = \lambda_1 \frac{(\theta+2-t)}{\theta(\theta+1)} H(t+1) - \lambda_2\left(1-\frac{3t}{\theta}+\frac{3t^2}{\theta^2}-\frac{t^3}{\theta^3}\right)H(t+1)$$

Considering the coefficient of $\frac{t^r}{r!}$, on both sides of the above expression, we obtain the following recurrence relation for moments of SPPL distribution.

$$\mu'_{(r+1)} = \left\{\lambda_1 \frac{(\theta+2)}{\theta(\theta+1)} + \frac{3r}{\theta} + \lambda_2\right\}\mu'_{(r)} - \left\{\lambda_1 \frac{r}{\theta(\theta+1)} + \frac{3r(r-1)}{\theta^2} + \lambda_2 \frac{3r}{\theta}\right\}\mu'_{(r-1)} +$$

$$\left\{\frac{r(r-1)(r-2)}{\theta^3} + \lambda_2 \frac{3r(r-1)}{\theta^2}\right\}\mu'_{(r-2)} - \lambda_2 \frac{r(r-1)(r-2)}{\theta^3}\mu'_{(r-3)} \qquad (5.3.3)$$

where $\mu'_{(r)}$ denotes the r^{th} order factorial moments of SPPL distribution. Explicit expression for the first four factorial moments may be obtained as

$$\mu'_{(1)} = \lambda_1 \frac{(\theta+2)}{\theta(\theta+1)} + \lambda_2$$

$$\mu'_{(2)} = \lambda^2{}_1 \frac{(\theta+2)^2}{\theta^2(\theta+1)^2} + 2\lambda_1 \frac{(\theta+3)}{\theta^2(\theta+1)} + 2\lambda_1\lambda_2 \frac{(\theta+2)}{\theta(\theta+1)} + \lambda^2{}_2$$

$$\mu'_{(3)} = \left\{\lambda_1 \frac{(\theta+2)}{\theta(\theta+1)} + \lambda_2 + \frac{6}{\theta}\right\}\mu'_{(2)} - \left\{\lambda_1 \frac{1}{\theta(\theta+1)} + \frac{6\lambda_2}{\theta} + \frac{6}{\theta^2}\right\}\mu'_{(1)} + 6\frac{\lambda_2}{\theta}$$

$$\mu'_{(4)} = \left\{\lambda_1 \frac{(\theta+2)}{\theta(\theta+1)} + \frac{9}{\theta} + \lambda_2\right\}\mu'_{(3)} - \left\{\frac{\lambda_1}{\theta(\theta+1)} + \frac{18}{\theta^2} + \lambda_2 \frac{9}{\theta}\right\}\mu'_{(2)} -$$

$$\left\{\frac{6}{\theta^3} + \lambda_2 \frac{18}{\theta^2}\right\}\mu'_{(1)} - \frac{6}{\theta^3}$$

Hence mean=$\lambda_1 \frac{(\theta+2)}{\theta(\theta+1)}+\lambda_2$ and variance=$\lambda_1 \frac{(\theta^2+4\theta+6)}{\theta^2(\theta+1)}+\lambda_2$

If $\lambda_2 \to 0$, the moments become same as those of Poisson-Poisson-Lindley distribution [see chapter 3].

5.4 Estimation of Parameters

A composite method has been used to estimate the parameters of SPPL distribution. The method is based on ratio of first two frequencies, sample mean and sample variance. We have

$$\bar{x}=\lambda_1 \frac{(\theta+2)}{\theta(\theta+1)}+\lambda_2 \tag{5.4.1}$$

and
$$S^2=\lambda_1 \frac{(\theta^2+4\theta+6)}{\theta^2(\theta+1)}+\lambda_2 \tag{5.4.2}$$

By equating the first two probabilities of equation (5.2.3) of SPPL distribution with $\frac{n_0}{N}$ and $\frac{n_1}{N}$, we obtain

$$\frac{n_1}{n_0}=\lambda_1 \frac{\theta^2(\theta+3)}{(\theta+1)^4}+\lambda_2 \tag{5.4.3}$$

By eliminating λ_1 and λ_2 between equations (5.4.1), (5.4.2) and (5.4.3), we have

$$\frac{2(\theta+1)^3(\theta+3)}{2\theta^4+9\theta^3+7\theta^2+2\theta}=\frac{S^2-\bar{x}}{\bar{x}-\frac{n_1}{n_0}} \tag{5.4.4}$$

which gives an estimate for θ either by graphically or by numerical solution, using Newton Raphson method i.e.

Let us consider $f(\theta)=\dfrac{2(\theta+3)(\theta+1)^3}{2\theta^4+9\theta^3+7\theta^2+2\theta}-K$

where $K=\dfrac{S^2-\bar{x}}{\bar{x}-\dfrac{f_1}{f_0}}$

Then the iteration formula for Newton Raphson method is

$$\hat{\theta}=\theta_0-\frac{f(\theta)}{f'(\theta)} \tag{5.4.5}$$

where θ_0 is the initial value and $\hat{\theta}$ is the estimated value of θ respectively. The initial guess value for starting the Newton Raphson method has to be selected by trial values, based on our assumptions. When the trial value closes to the estimated value, the method will always be convergent.

After getting the estimate of θ i.e. $\hat{\theta}$ from equation (5.4.5), the estimate of λ_1 and λ_2 may be obtained as

$$\hat{\lambda}_1=\frac{(S^2-\bar{x})\hat{\theta}^2(\hat{\theta}+1)}{2(\hat{\theta}+3)} \tag{5.4.6}$$

$$\hat{\lambda}_2=\bar{x}-\frac{\hat{\lambda}_1(\hat{\theta}+2)}{\hat{\theta}(\hat{\theta}+1)} \tag{5.4.7}$$

5.5 Some Special cases of SPPL distribution

The limiting forms of the SPPL distribution as the parameters take particular values are given in Table 5.1.

Table 5.1

Sl.no.	Parameter values	Distribution and its pgf
1	$\theta = 0$, $\lambda_1 = 0$	Poisson, $e^{\lambda_2(t-1)}$
2	$\lambda_2 = 0$	Poisson-Poisson-Lindley, $e^{\lambda_1\left\{\frac{\theta^2(\theta+2-t)}{(\theta+1)(\theta+1-t)^2}-1\right\}}$

5.6. Goodness of Fit

To illustrate the applications of the SPPL distribution, firstly, in Table 5.2 we consider the data on the number of accidents sustained by a group of 708 bus drivers over a period of three years for which Neyman Type A and 'Short' distributions were fitted by Kemp (1967). When SPPL distribution is applied to these data it provides a good fit with χ^2 value of 2.445. Using the equations (5.4.5), (5.4.6) and (5.4.7), we get $\hat{\theta}$ =8.0110, $\hat{\lambda}_1$ =30.171 and $\hat{\lambda}_2$= -1.892.

Secondly, in Table 5.2, we have considered the data on number of accidents to 647 women on high explosive shells during 5 weeks period [data from Greenwood & Yule, (1920)]. Here the original data together with the expected frequencies of SPPL, PPL [Borah & Deka Nath (2001)] and Negative binomial [Plunket & Jain (1975)] distribution are considered.. Using equations (5.4.5), (5.4.6) and (5.4.7), we get $\hat{\theta}$ =5.3066, $\hat{\lambda}_1$=2.4117 and $\hat{\lambda}_2$=-0.061. From Table 5.3, it is seen that SPPL distribution provides a better fit to this data as compared to the other distributions.

Considering Adelstein (1952) data on number of accidents (home injuries) of 122 experienced men in 11 years period where Adelstein had concluded that the Poisson distribution fits the first sets of data but did not fit the second sets. In Table 5.4, when the SPPL distribution is applied to those sets it provides a good fit like GPD [Consul (1989)] model in all the cases. Here we get $\hat{\theta}$ =4.2305, $\hat{\lambda}_1$=0.4401 and $\hat{\lambda}_2$ =0.4171 for the first set and $\hat{\theta}$=4.0447, $\hat{\lambda}_1$ =3.359 and $\hat{\lambda}_2$=-0.0115 for the second set. When the estimated value of θ are divided by the respective number of years (6 & 11years), the average values of $\hat{\theta}$ for these sets become 0.705 & 0.3677 respectively. These values do indicate that the average natural rate for home injuries does decrease with experience.

In Table 5.5, we have considered accident data for experienced shunting men over 11 years. For which Adelstein (1952) had used the Poisson distribution and negative binomial distribution. The SPPL distribution is fitted to this data and it is found that the calculated value of χ^2 for the SPPL distribution is much less than its significant values. In Table 5.5, we get $\hat{\theta}$ =6.7819, $\hat{\lambda}_1$=6.759 and $\hat{\lambda}_2$=0.1457 for the first set and $\hat{\theta}$ =1.7748, $\hat{\lambda}_1$=0.2090 and $\hat{\lambda}_2$=0.8069 for the second set. The values of $\hat{\theta}$ for the second group indicates that due to experienced, this group has a less natural chance of making accidents.

It is apparent from the results of the following tables that the SPPL distribution can be applied very successfully in case of accident data. In all the cases the SPPL distribution provides a good fit.

Table 5.2 Numbers of drivers sustaining accidents over three year period [Cresswell & Froggatt's Table 5.4, (1963)] with expected frequencies based on SPPL , Neyman –Type A [Kemp (1967)] and Short distribution [Kemp (1967)].

No. of accidents	observed frequency	Expected frequency			
		SPPL	NTA	Short(MM)	Short(ML)
0	117	118.588	116.69	110.38	116.88
1	157	159.130	162.04	169.70	160.43
2	158	153.191	153.12	156.02	153.64
3	115	115.631	115.26	113.90	116.05
4	78	74.94	74.58	72.54	75.13
5	44	43.182	43.13	41.90	43.29
6	21	22.687	22.83	22.45	22.76
7	7	11.095	11.25	11.31	11.09
8	6	5.041	5.21	5.42	5.07
9	1	2.177	2.29	2.49	2.20
10	3	0.895	0.96	1.10	0.91
11	1	1.493	0.39	0.43	0.35
Total	708	708.00	707.75	707.68	707.81
χ^2		2.445	2.64	3.78	2.56

Note. MM: Method of moments
ML: Method of maximum likelihood.

Table 5.3 Comparison of Observed Frequencies for Accidents to 647 Women on High Explosive Shells during 5 weeks with Expected Frequencies of SPPL, PPL[Borah & Deka Nath (2001)] and Negative Bionomial(NB) distribution[Plunket & Jain (1975)]. [Data from Greenwood and Yule (1920)].

No. of accidents	Observed frequency	Expected frequency		
		SPPL	PPL	NB
0	447	445.959	442.52	445.89
1	132	131.692	137.79	134.90
2	43	47.698	46.57	44.00
3	21	15.218	14.57	14.69
4	3	4.124	4.32	4.94
≥5	2	2.309	1.70	2.56
Total	647	647.00	647.00	647.00
χ^2		2.981	3.57	3.6315

Table 5.4 Comparison of Observed Frequencies for Home Injuries of 122 Experienced Men during 5 years (1937-1942) and 11 years (1937-1947) with Expected SPPL & GPD [Consul (1989)] Frequencies. [Data from Adelstein (1952)]

No. of Injuries	1937-1942			1937-1947		
	Obs.	SPPL	GPD	Obs.	SPPL	GPD
0	73	72.952	73.23	58	57.045	57.22
1	36	35.977	35.32	34	33.441	34.41
2	10	10.079	10.41	14	17.521	16.64
3	2	2.313	3.04	8	8.161	7.59
4	1	0.679		6	3.404	6.14
5	-			2	2.428	
Total	122	122.00	122.00	122	122.00	122.00
χ^2		0.00068	2.62		1.542	1.09

Note: Obs.: Observed frequency,
GPD: Generalized Poisson distribution

Table 5.5 Comparison of Observed Frequencies of Accidents of 122 Experienced Shunting Men over 11 years (1937-1947) with Expected SPPL & GPD [Consul (1989)] Frequencies. [Data from Adelstein (1952)].

No. of	1937-1942			1937-1947		
Injuries	Obs	SPPL	GPD	Obs.	SPPL	GPD
0	40	40.141	39.98	50	49.642	51.48
1	39	39.138	39.49	43	39.671	39.57
2	26	23.794	23.76	17	19.945	19.28
3	8	11.459	11.32	9	6.731	7.67
4	6	4.749	4.70	2	2.074	4.00
5	2	1.765	2.75	0	0.579	
6	1	0.954		1	0.808	
Total	122	122.00	122.00	122	122.00	122.00
χ^2		1.648	1.57		1.15	1.09

CHAPTER 6

- A CLASS OF HERMITE TYPE LAGRANGIAN DISTRIBUTIONS

Chapter 6

A Class of Hermite Type Lagrangian Distributions

6.1 Introduction

Lagrangian expansion for the derivation of expressions for probabilities of certain discrete distributions were used by Consul and Shenton (1972,1973,1975), Mohanty (1966), Consul and Jain (1973) and their co workers. The nature of the generalization process for these distributions was clarified in two important papers by Consul and Shenton (1972,1973) and also by Consul (1983). Consul's (1989) book on generalized Poisson distributions offers a systematic study of the Lagrangian Poisson distribution. This book also focuses on applications of the generalized Poisson model to various areas with actual data sets. Consul and Famoye (1996) studied Lagrangian Katz family of distributions. The parameters of this distribution were estimated by various method of estimation and some of the applications of the Lagrangian Katz distribution were studied.

If g(t) and f(t) be two probability generating functions (pgf) defined on non-negative integers, such that $g(0) \neq 0$, then the pgf for the general distribution formed from g(t) and f(t) where considering the transformation t=u.g(t) is given by [see Consul and Shenton (1972)]

$$f(t) = f(0) + \sum_{s=1}^{\alpha} \frac{u^s d^{s-1}}{s!\, dt^{s-1}} \left\{ (g(t))^s . f'(t) \right\} |_{t=0} \qquad (6.1.1)$$

Thus the probability mass function (pmf) for the general Lagrangian probability distribution is given by

$$P_r[X = x] = \frac{1}{x!} \frac{d^{x-1}}{dt^{x-1}} \left\{ (g(t))^x . f'(t) \right\} |_{t=0} \qquad x=1,2.3,.... \qquad (6.1.2)$$

$$P_r[X = 0] = f(0)$$

Equation (6.1.2) is also known as Lagrangian distribution of type-I (LD-I) (according Janardan and Rao's terminology).

Using Lagrangian expansion of 2nd kind, Janardan and Rao (1983) investigated a new class of discrete distribution call Lagrangian distribution of type -II (LD-II) with pmf

$$P(X = x) = \begin{cases} \frac{\{1 - g'(1)\}}{x!} \left[\frac{\delta^x}{\delta z^x} \left\{ (g(z))^x f(z) \right\} \right]_{z=0}, & x = 0,1,2,.... \\ 0, otherwise & \end{cases} \qquad (6.1.3)$$

The Hermite distribution was formally introduced by Kemp and Kemp (1966) and was applied in the field of biological sciences, physical sciences and operation research. The pgf of this distribution is

$$\exp\left[\alpha_1 (t - 1) + \alpha_2 (t^2 - 1)\right]$$

The probabilities of this distribution can be conveniently expressed in terms of modified Hermite polynomials.

The motivation of investigation behind this chapter is to derive the Hermite type Lagrangian distribution, since Hermite distribution has a wide application in various fields of experimentation. The basic Lagrangian Hermite distribution (LHD) is investigated. The parameters of this distribution are estimated by the method of moments and ratio of first frequency and mean. Considering different values of f(t) and g(t) in equation (6.1.2) and (6.1.3), different general Lagrangian Hermite type distributions are generated. Borah and Deka Nath (2000) studied Lagragian Hermite type distributions and fitted this distribution to some well known data sets for empirical comparison.

6.2 Basic Lagrangian Hermite Distribution (LHD)

The pmf of basic Lagrangian distribution is given as

$$P_r[X=x]=\begin{cases}\frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\{g(t)\}^x\Big|_{t=0}, & x=1,2,3,\ldots \\ 0, & otherwise\end{cases} \tag{6.2.1}$$

where g(t) is the pgf defined on some or all non-negative integers, such that $g(0)\neq 0$. In this case

$$g(t)=\exp[\alpha_1(t-1)+\alpha_2(t^2-1)] \tag{6.2.2}$$

Thus the pmf of basic LHD may be written as

$$P_r[X=x]=\begin{cases}\exp[-x(\alpha_1+\alpha_2)]\sum_{j=0}^{[(x-1)/2]}\frac{\alpha_1^{x-1-2j}\alpha_2^j x^{x-1-j}}{(x-2j)!\,j!}, & x=1,2,3,\ldots, \\ 0, & otherwise\end{cases} \tag{6.2.3}$$

where $[(x-1)/2]$ denotes the integer part of (x-1)/2

(a) Cumulants of the Basic LHD

The cumulants of basic Lagrangian distribution may be investigated by using Consul and Shenton (1975) general formula. According to this , if G_i be the i[th] cumulant of the distribution with pgf, g(t) then the first four cumulants k_i, i=1,2,3,4 of basic Lagrangian distribution can be written as

$$k_1=\frac{1}{1-G_1},$$

$$k_2=\frac{G_2}{(1-G_1)^3},$$

$$k_3=\frac{G_3}{(1-G_1)^4}+\frac{3G_2^{\ 2}}{(1-G_1)^5},$$

$$k_4=\frac{G_4}{(1-G_1)^5}+\frac{10G_3G_2}{(1-G_1)^6}+\frac{15G_3^{\ 2}}{(1-G_1)^7}$$ [see Consul and Shenton, 1975] .

In case of Hermite distribution, we have the first four cumulants as

$$G_1=\alpha_1+2\alpha_2 \qquad (6.2.4)$$

$$G_2=\alpha_1+4\alpha_2 \qquad (6.2.5)$$

$$G_3=\alpha_1+8\alpha_2 \qquad (6.2.6)$$

$$G_4=\alpha_1+16\alpha_2 \qquad (6.2.7)$$

Thus the first four cumulants of basic LHD may be written as

$$k_1=\frac{1}{(1-\alpha_1-2\alpha_2)}, \qquad (6.2.8)$$

$$k_2 = \frac{\alpha_1 + 4\alpha_2}{(1-\alpha_1-2\alpha_2)^3} \tag{6.2.9}$$

$$k_3 = \frac{\alpha_1 + 8\alpha_2}{(1-\alpha_1-2\alpha_2)^4} + \frac{3(\alpha_1+4\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^5}, \tag{6.2.10}$$

$$k_4 = \frac{\alpha_1 + 16\alpha_2}{(1-\alpha_1-2\alpha_2)^5} + \frac{10(\alpha_1+8\alpha_2)(\alpha_1+4\alpha_2)}{(1-\alpha_1-2\alpha)^6} + \frac{15(\alpha_1+8\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^7}. \tag{6.2.11}$$

(b) Estimation of Parameters

Method of moments and ratio of first frequency with mean are used to estimate the parameters of basic LHD.

(i) Method of moments

The mean and variance of basic LHD as given in equation (6.2.8) and (6.2.9), may be written as

$$\bar{x} = \frac{1}{(1-\alpha_1-2\alpha_2)} \tag{6.2.12}$$

$$m_2 = \frac{\alpha_1 + 4\alpha_2}{(1-\alpha_1-2\alpha_2)^3} \tag{6.2.13}$$

where $\bar{x}$ is the sample mean and m_2 is the sample variance.

By eliminating α_1 between equation (6.2.12) and (6.2.13), we obtain

$$\hat{\alpha}_2 = \frac{1}{2}\left[\frac{m_2}{\bar{x}^3} + \frac{1}{\bar{x}} - 1\right] \tag{6.2.14}$$

Substituting the value of $\hat{\alpha}_2$ in equation (6.2.12), we may get

$$\hat{\alpha}_1 = 2 - \frac{m_2}{\bar{x}^3} - \frac{2}{\bar{x}} \tag{6.2.15}$$

(ii) Ratio of first frequency and mean

By equating the first probability of basic LHD with $\frac{n_1}{N}$, we obtain

$$\frac{n_1}{N} = \exp(-\alpha_1 - \alpha_2) \tag{6.2.16}$$

By eliminating α_1 between equation (6.2.12) and (6.2.16), we may obtain

$$\hat{\alpha}_2 = 1 - \frac{1}{\bar{x}} + \log \frac{n_1}{N} \tag{6.2.17}$$

and

$$\hat{\alpha}_1 = \frac{1}{\bar{x}} - 2\log \frac{n_1}{N} - 1 \tag{6.2.18}$$

(c) Fitting of Basic LHD

For the application of basic LHD, we have considered the example of numbers of papers published per author in the review of applied entomology [data by Kendall (1961)] in Table 6.1, for which geometric distribution (GD), logarithmic series distribution (LSD) were fitted by Williams (1944) and generalized logarithmic series distribution (GLSD) was fitted by Jain (1975).

From the Table 6.1, we have, the sample mean $\bar{x} = 1.5508475$ and sample variance $m_2 = 1.1405050$. By using method of moments we have $\hat{\alpha}_1 = 0.404616$ and $\hat{\alpha}_2 = -0.024712$ and by using ratio of first frequency and mean we have $\hat{\alpha}_1 = 0.3803$ and $\hat{\alpha}_2 = -0.012$. From the Table 6.1, it is observed that the LHD model (by using both the method of estimation) gives better fit than GD, LSD and GLSD as judged by the values of χ^2.

Table 6.1 Fitting of no. of papers per author by LHD, GLSD, LSD and GD. Publication in the review of applied entomology. Vol. 24, 1936 (2379 papers by 1534 authors)

No.of papers per author	Observed frequency	LHD (MM)	LHD (FM)	GLSD Jain(1975)	LSD Williams (1944)	GD Williams (1944)
1	1062	1049.14	1062.06	1052.72	1046.05	989.10
2	263	290.33	279.59	287.52	293.05	351.30
3	120	108.38	104.02	107.10	109.46	124.80
4	50	45.86	45.95	45.10	45.99	44.33
5	22	20.78	21.33	20.83	20.61	15.75
6	7	13.96	10.33	10.00	9.62	5.59
7	6	3.09	5.45	4.97	4.62	1.99
8	2	1.49	2.30	2.53	2.26	0.71
9	0	0.74	1.23	1.31	1.12	0.25
10	1	0.15	1.19	0.70	0.53	0.09
11	1	0.09	0.55	1.81	0.66	0.09
Total	1534	1534.00	1534.00	1534.00	1534.00	1534.00
χ^2		4.74	4.94	5.14	5.56	46.39
Parameter estimates		$\hat{\alpha}_1 = 0.4046$ $\hat{\alpha}_2 = -0.024$	0.3803 -0.012			

Note: MM: Method of moments, FM: Ratio of mean and first frequency

6.3 General Lagrangian Hermite Poisson Distribution

Considering different values of g(t) and f(t) in equation (6.1.2) and (6.1.3), different LHD of type I and type II may be obtained. Let

$$g(t)= \exp[\alpha_1(t-1)+\alpha_2(t^2-1)]$$

$$f(t)=\exp[\theta(t-1)]$$

where g(t) be the pgf of Hermite distribution and f(t) be the pgf of Poisson distribution respectively. Hence the pmf of Lagrangian Hermite Poisson distribution of type I (LHPD-I) may be written as

$$P_r(X=x)=\begin{cases}\theta\exp\{\theta+x(\alpha_1+\alpha_2)\}\sum_{j=0}^{\left[\frac{x-1}{2}\right]}\frac{(\theta+x\alpha_1)^{x-1}\alpha_2{}^j x^j}{(x-2j)!\,j!}, & x=1,2,3,\ldots\\ \exp[-\theta], & x=0\end{cases} \quad (6.3.1)$$

where $\theta>0$.

Similarly, considering equation (6.1.3), the pmf of LHPD-II may be written as

$$P_r(X=x)=\begin{cases}A\exp\{\theta+x(\alpha_1+\alpha_2)\}\sum_{j=0}^{\left[\frac{x}{2}\right]}\frac{(\theta+x\alpha_1)^{x-2j}\alpha_2^j x^j}{(x-2j)!\,j!}, & x=0,1,2,\ldots\\ 0, & otherwise.\end{cases} \quad (6.3.2)$$

where A={1-($\alpha_1+2\alpha_2$)}

(a) Cumulants of General LHPD

According to Consul and Shenton (1975), if F_r be the r[th] cumulants for the pgf f(t) as a function of t, and if k_r be the r[th] cumulants for the basic Lagrangian distribution obtained from g(t) then the cumulants of general Lagrangian distribution may be written as

$$D_1=F_1k_1,$$

$$D_2=F_1k_2+F_2k_1^2,$$

$$D_3=F_1k_3+3F_2k_1k_2+F_3k_1^3,$$

$$D_4=F_1k_4+3F_2k_2^2+4F_2k_1k_3+6F_3k_1^2k_2+F_4k_1^4.$$

(See Consul and Shenton 1975).

Here D_1, D_2, D_3 and D_4 are the first four cumulants of general Lagrangian distribution. k_1, k_2, k_3 and k_4 are given in equation (6.2.8), (6.2.9), (6.2.10) and (6.2.11) respectively. Thus the first three cumulants of general LHPD may be written as

$$D_1 = \frac{\theta}{(1-\alpha_1-2\alpha_2)}, \tag{6.3.3}$$

$$D_2 = \frac{\theta(\alpha_1+4\alpha_2)}{(1-\alpha_1-2\alpha_2)^3} + \frac{\theta}{(1-\alpha \ -2\alpha_2)^2}, \tag{6.3.4}$$

$$D_3 = \frac{3\theta(\alpha_1+4\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^5} + \frac{4\theta(\alpha_1+5\alpha_2)}{(1-\alpha_1-2\alpha_2)^4} + \frac{\theta}{(1-\alpha_1-2\alpha_2)^3}. \tag{6.3.5}$$

(b) Estimation of Parameters

The parameters of LHPD-I may be estimated from the ratio of first two frequencies and the mean. By equating the first and second probabilities of LHPD-I with $\frac{n_0}{N}$ and $\frac{n_1}{N}$, we obtain

$$\hat{\theta} = -\log\frac{n_0}{N}, \tag{6.3.6}$$

and $$\log\frac{n_1}{N} = \log\theta - (\theta+\alpha_1+\alpha_2) \tag{6.3.7}$$

Also from equation (6.3.3), we have

$$\bar{x} = \frac{\theta}{1-\alpha_1-2\alpha_2}, \tag{6.3.8}$$

Hence from equation (6.3.6), (6.3.7) and (6.3.8), we may estimate α_1 and α_2 as

$$\hat{\alpha}_2 = 1 - \frac{\hat{\theta}}{\bar{x}} + \log\frac{n_1}{N} - \log\hat{\theta} + \hat{\theta} \tag{6.3.9}$$

and $$\hat{\alpha}_1 = 1 - \frac{\hat{\theta}}{\bar{x}} - 2\hat{\alpha}_2 \qquad (6.3.10)$$

(c) Fitting of LHPD-I

In Table 6.2, we have considered Adelstein (1952) data on number of accidents (home injuries) of 122 experienced men in 11 years period where Adelstein concluded that the Poisson distribution gave good fit to the first sets of data but did not fit the second and third sets..

When LHPD-I is fitted to these data sets it gives good fit to all the cases. From the following table it is also observed that the LHPD-I distribution gives better fit than GPD (Consul, 1989). In Table 6.2, we get $\hat{\theta}$ =0.5135, $\hat{\alpha}_1$ =0.0103 and $\hat{\alpha}_2$ =0.0304 for the 1st set and $\hat{\theta}$ =0.3267, $\hat{\alpha}_1$ =-0.1799 and $\hat{\alpha}_2$ = 0.648 for the 2nd set and $\hat{\theta}$ =0.7435, $\hat{\alpha}_1$ =0.006 and $\hat{\alpha}_2$=0.2316 for 3rd set.

Thus this chapter has defined and studied a class of Hermite type Lagrangian probability distribution, by well known Lagrange's expansion, with application to various fields of data . The fitting of LHPD-II will be invistigated latter on.

Table 6.2: Comparison of Observed Frequencies for Home Injuries of 122 Expected Men during 11 years (1937-1947) with the Expected LHPD-I and GPD [Consul (1989)] Frequencies.

Number of injuries	1937-1942			1943-1947			1937-1947		
	Observed	LHPD	GPD	Observed	LHPD	GPD	Observed	LHPD	GPD
0	73	72.99	73.23	88	87.99	86.77	58	57.99	57.22
1	36	35.99	35.32	18	17.99	22.85	34	33.99	34.41
2	10	9.93	10.41	11	9.14	7.56	14	16.17	16.64
3	2	2.36	3.04	4	4.79	2.82	8	7.42	7.59
4	1	0.73	-----	1	2.09	2.00	6	3.41	6.14
5							2	3.02	------
Total and Estimates	122	$\hat{\theta}$=0.5135		122	0.3276		122	0.7435	
		$\hat{\alpha}_1$=0.0103			-0.1799			0.006	
		$\hat{\alpha}_2$=0.0304			0.648			0.2316	
		$\chi^2 = 0.15$	2.62		0.00002	0.124		0.719	1.09

CHAPTER 7

- A CLASS OF CHARLIER TYPE LAGRANGIAN PROBABILITY DISTRIBUTIONS

Chapter 7

A class of Charlier Type Lagrangian Distributions

7.1 Introduction

Doetsch (1933), Meixner (1934,1938) and Berg (1985) investigated Charlier polynomials which was defined by the generating function

$$e^{z}(1-\beta z)^{-\lambda}$$

Jain and Gupta (1975) defined the generalized Charlier polynomial by the generating function

$$e^{z\alpha}(1-\beta z^{m})^{-\lambda}$$

Medhi and Borah (1986) studied the generalized four parameter Charlier distribution with pgf

$$H(z)=e^{-\alpha}(1-\gamma-\beta)^{\lambda}\left(1-\gamma z-\beta z^{m}\right)^{-\lambda}e^{\alpha z},\quad \alpha,\beta,\gamma,\lambda\geq 0 \text{ and } \text{m}=1,2,\ldots$$

They studied some properties of this distribution including recurrence relations for the probability mass function as well as for the moments and cumulants of the distribution. Negative binomial, Gegenbauer, generalized Gegenbauer, Charlier and generalized Charlier distributions were the limiting distribution of this four parameter Charlier distribution. Medhi and Borah (1986) also discussed the methods for fitting of these four parameter Charlier distribution.

Using Lagrange expansion to this Charlier distribution, Borah and Begum (1997) studied only the probabilistic structures of Lagrangian Charlier distribution of type-I (LCD-I) and type-II (LCD-II). Deka Nath and Borah (2001) studied the basic LCD and fitted this distribution to a data set for empirical comparison.

The objective of this chapter is two folded. Firstly, we have to investigate the pmf of basic LCD in a simpler form than the earlier one, which is also easy to handle on computer. Secondly, considering $f(t) = e^{\theta(t-1)}$ in equation (6.1.2) and (6.1.3), Lagrangian Charlier Poisson distribution of type I and type II (LCPD-I & LCPD-II) are also derived. The cumulants of the distributions are investigated. For fitting of basic LCD, some methods of estimation of the parameters are suggested. The basic LCD has been fitted to some data for which logarithmic series, geometric, generalized logarithmic series and basic Lagragian Hermite distribution were fitted. It has been found that the basic LCD gives surprisingly a better fit than the other distributions. The LCPD-I is also fitted to some data sets for which generalized Poisson distribution (GPD) was fitted by Consul (1989).

7.2 Basic Lagrangian Charlier Distribution (LCD)

In this chapter we have considered the pgf of three parameter Charlier distribution as

$$g(z)=e^{-\alpha}(1-\beta)^{\lambda}e^{\alpha z}(1-\beta z)^{-\lambda} \tag{7.2.1}$$

Thus the pmf of basic LCD may be written as

$$P(X=x)=\begin{cases}\frac{e^{-x\alpha}(1-\beta)^{\lambda x}}{x!}\left\{\sum_{j=0}^{k}\binom{k}{j}(\alpha x)^{k-j}\beta^{j}(\lambda x)_{(j)}\right\}, & x=1,2,\ldots \\ 0, & otherwise.\end{cases} \tag{7.2.2}$$

where k=x-1, $\alpha>0$, $\beta<1$, $\lambda>0$.

Equation (7.2.2) may also be written as

$$P(X=x)=\frac{e^{\alpha x}(1-\beta)^{\lambda x}(\alpha x)^{x-1}}{x!}{}_2F_0(1-x,\lambda x,-\beta/\alpha x),\ x\geq 1 \tag{7.2.3}$$

[see Borah and Begum (1997)]

(a) Cumulants of the Basic LCD

Using Consul and Shenton (1975) general formula, the cumulants of basic LCD are investigated. Let G_i be the i^{th} cumulants of the Charlier distribution with pgf g(z), then we have

$$G_1=\alpha+\frac{\lambda\beta}{(1-\beta)} \tag{7.2.4}$$

$$G_2=\alpha+\frac{\lambda\beta}{(1-\beta)^2} \tag{7.2.5}$$

$$G_3=\alpha+\frac{\lambda\beta(1+\beta)}{(1-\beta)^3} \tag{7.2.6}$$

$$G_4=\alpha+\frac{\lambda\beta}{(1-\beta)^4} \tag{7.2.7}$$

then the first four cumulants k_i ,i=1,2,3,4 for basic LCD may be obtained by the relations as shown in chapter 6.

$$k_1=\frac{(1-\beta)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}}, \qquad (7.2.8)$$

$$k_2=\frac{(1-\beta)\{\alpha(1-\beta)^2+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3}, \qquad (7.2.9)$$

$$k_3=\frac{\alpha(1-\beta)^4+\lambda\beta(1-\beta^2)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^4}+\frac{3(1-\beta)\{\alpha(1-\beta)^4+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^5}, \qquad (7.2.10)$$

$$k_4=\frac{(1-\beta)}{[(1-\alpha)(1-\beta)-\lambda\beta]^5}\left[\{\alpha(1-\beta)^4+\lambda\beta\}+15\frac{\{\alpha(1-\beta)^2+\lambda\beta\}^3}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^2}\right.$$

$$\left.+10\frac{\{\alpha(1-\beta)^3+\lambda\beta(1+\beta)\}\{\alpha(1-\beta)^2+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}}\right] \qquad (7.2.11)$$

(b) Estimation of Parameters

The parameters of basic LCD can be estimated in terms of its cumulants. A composite method has been used to estimate the parameters. The method is based on ratio of first two moments and first frequency.

By equating the first probability of basic LCD with $\frac{n_1}{N}$, we obtain

$$\alpha=\lambda\log(1-\beta)-\log(\frac{n_1}{N}) \qquad (7.2.12)$$

By equating the mean and variance of the sample to the population value of mean and variance of basic LCD , given in equation (7.2.4) and (7.2.5) , we get

$$\bar{x}=\frac{(1-\beta)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}} \qquad (7.2.13)$$

$$m_2=\frac{(1-\beta)\{\alpha(1-\beta)^2+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3} \qquad (7.2.14)$$

By eliminating α and λ between equation (7.2.12), (7.2.13) and (7.2.14), we obtain

$$(\frac{1}{\beta}-1)^2\log(1-\beta)+\frac{1}{\beta}=\frac{m_2+\bar{x}^3\log(\frac{n_1}{N})}{m_2+\bar{x}^2-\bar{x}^3} \qquad (7.2.15)$$

The equation (7.2.15) may give an estimate for β either by graphically or by using Newton Raphson method.

After getting the estimate $\hat{\beta}$ of β from equation (7.2.15), the estimates of λ and α may be obtained from the following equations.

$$\hat{\lambda}=\frac{\log(\frac{n_1}{N})-\frac{1}{\bar{x}}+1}{\log(1-\hat{\beta})+\frac{\hat{\beta}}{(1-\hat{\beta})}}, \qquad (7.2.16)$$

and $$\hat{\alpha}=1-\frac{1}{\bar{x}}-\frac{\hat{\lambda}\hat{\beta}}{(1-\hat{\beta})} \qquad (7.2.17)$$

(c) Fitting of Basic LCD

For the application of basic LCD, we consider the example of number of papers published per author for which geometric distribution (GD) and logarithmic series distribution (LSD) were fitted by Williams (1944) and by generalized logarithmic series distribution (GLSD) by Jain (1975). The comparison of observed and expected frequencies among LCD, GLSD, LSD and GD are given in the following Table 7.1.

For the data in Table 7.1, we have the sample mean $\bar{x}$=1.5508475 and central moments m_2=1.1405050. Solving the equation (7.2.15) by Newton Raphson method, we get $\hat{\beta} = 0.6113$. Substituting the values of $\hat{\beta}$ in equation (7.2.16) and (7.2.17), we get $\hat{\lambda}$=-0.0199 and $\hat{\alpha}$=0.3866. It is clear from Table 7.1 that the expected basic LCD frequencies are much closer to the observed frequencies than obtained by geometric, logarithmic and generalized logarithmic distributions. Thus the LCD model better describes the pattern of the frequency distribution of number of paper per author.

Table 7.1. Fitting of number papers per author by LCD, GLSD, LSD and GD. Publication in the review of applied entomology, Vol. 24, 1936 (2379 papers by 1534 authors).

No.of papers per author	Observed frequency	LCD $\hat{\alpha} = 0.3866$ $\hat{\lambda} = -0.0199$	LHD (FM)	GLSD Jain (1975)	LSD Williams (1944)	GD Williams (1944)
1	1062	1061.90	1062.06	1052.72	1046.05	989.10
2	263	275.02	279.59	287.52	293.05	351.30
3	120	105.14	104.02	107.10	109.46	124.80
4	50	46.08	45.95	45.10	45.99	44.33
5	22	22.53	21.33	20.83	20.61	15.75
6	7	11.41	10.33	10.00	9.62	5.59
7	6	5.76	5.45	4.97	4.62	1.99
8	2	3.15	2.30	2.53	2.26	0.71
9	0	1.47	1.23	1.31	1.12	0.25
10	1	0.43	1.19	0.70	0.53	0.09
11	1	0.19	0.55	1.81	0.66	0 09
Total	1534	1534.00	1534.00	1 534.00	1534.00	1534.00
χ^2		4.66	4.94	5.14	5.56	46.39

Note. FM: Ratio of mean and first frequency.

7.3 General Lagrangian Charlier type Distributions

Considering $g(z)=e^{-\alpha}(1-\beta)^{\lambda}e^{\alpha z}(1-\beta z)^{-\lambda}$

and $f(z)=e^{\theta(z-1)}$

in equation (6.1.1), the pmf of Lagrangian Charlier Poission distribution of type I (LCPD-I) may be written as

$$P(X=x)=\frac{e^{-(x\alpha+\theta)}(1-\beta)^{\lambda x}\theta}{x!}\left\{\sum_{j=0}^{k}\binom{k}{j}(\alpha x+\theta)^{k-j}\beta^{j}(\lambda x)_{(j)}\right\}, \quad \text{for } x=1,2,\ldots \tag{7.3.1}$$

and $P(X=0)=e^{-\theta}$

where k=x-1 , $\alpha,\lambda,\theta>0$ and $\beta<1$.

Similarly, considering equation (6.1.2), the pmf of LCPD-II may be written as

$$P(X=x)=A\frac{e^{-(x\alpha+\theta)}(1-\beta)^{\lambda x}}{x!}\left\{\sum_{j=0}^{x}\binom{x}{j}(\alpha x+\theta)^{x-j}\beta^{j}(\lambda x)_{(j)}\right\}, \quad \text{for } x=0,1,\ldots \tag{7.3.2}$$

where $A=1-\left\{\alpha+\frac{\lambda\beta}{(1-\beta)}\right\}$

(a) Cumulants of General Lagrangian Charlier Poisson Distribution(LCPD)

The following cumulants of general LCPD model can be derived by using Consul and Shanton (1975) formula as shown in chapter 6.

$$D_1=\frac{\theta(1-\beta)}{(1-\alpha)(1-\beta)-\lambda\beta}, \tag{7.3.3}$$

$$D_2=\frac{\theta(1-\beta)\{(1-\beta)^2+\lambda\beta^2\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3}, \tag{7.3.4}$$

$$D_3 = \frac{\theta(1-\beta)^3}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3}\left[\frac{4\alpha(1-\beta)+\frac{\lambda\beta}{(1-\beta)}(2+\beta)}{(1-\alpha)(1-\beta)-\lambda\beta}+\frac{3\left\{\alpha(1-\beta)^2+\frac{\lambda\beta}{(1-\beta)^2}\right\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^2}+1\right]. \qquad (7.3.5)$$

(b)Estimation of parameters

The parameters of LCPD-I can be estimated from the ratio of first two frequencies and by using the mean and variance of the distribution. By equating first two probabilities with $\frac{n_0}{N}$ and $\frac{n_1}{N}$ respectively, we obtain

$$\frac{n_0}{N} = e^{-\theta} \Rightarrow \hat{\theta} = -\log\left(\frac{n_0}{N}\right) \qquad (7.3.6)$$

and $$\frac{n_1}{N} = e^{-(\alpha+\hat{\theta})}(1-\beta)^{\lambda}\hat{\theta} \qquad (7.3.7)$$

$$\Rightarrow \quad \alpha = \lambda\log(1-\beta) - \log\left(\frac{n_1}{n_0}\right) + \log\hat{\theta} \qquad (7.3.8)$$

From equation (7.3.3), (7.3.4) we may get the following equations as

$$\bar{x} = \frac{\theta(1-\beta)}{(1-\alpha)(1-\beta)-\lambda\beta} \qquad (7.3.9)$$

and $$m_2 = \frac{\bar{x}}{\theta^2}\left\{1+\frac{\lambda\beta^2}{(1-\beta)^2}\right\} \qquad (7.3.10)$$

The estimate of λ in terms of θ and β may be obtained from equation (7.3.10) as $$\lambda = \left(\frac{m_2\theta^2}{\bar{x}^3} - 1\right)\frac{(1-\beta)^2}{\beta^2} \tag{7.3.11}$$

Again from equation (7.3.9) and (7.3.10), we obtain

$$\alpha = 1 - \frac{\theta}{\bar{x}} - \left(\frac{m_2\theta^2}{\bar{x}^3} - 1\right)\frac{(1-\beta)}{\beta} \tag{7.3.12}$$

By eliminating α between equation (7.3.8) and (7.3.12), we get

$$\lambda = \frac{1 - \frac{\theta}{\bar{x}} - \left(\frac{m_2\theta^2}{\bar{x}^3} - 1\right)\frac{(1-\beta)}{\beta} - \log\theta + \log(n_1/n_0)}{\log(1-\beta)} \tag{7.3.13}$$

Hence from equation (7.3.11) and (7.3.13), we get

$$\frac{(1-\beta)}{\beta}\left\{1 + \frac{(1-\beta)}{\beta}\log(1-\beta)\right\} = \frac{1 - \frac{\hat{\theta}}{\bar{x}} + \log(n_1/n_0) - \log\hat{\theta}}{\left(\frac{m_2\hat{\theta}^2}{\bar{x}^3} - 1\right)} \tag{7.3.14}$$

which gives an estimate for β either by graphically or by numerical solution using Newton-Raphson method. On getting the estimate $\hat{\theta}$ of θ from (7.3.6) and $\hat{\beta}$ of β from (7.3.14), the estimates of λ and α may be obtained by

$$\hat{\lambda} = \left(\frac{m_2\hat{\theta}^2}{\bar{x}^3} - 1\right)\frac{(1-\hat{\beta})^2}{\hat{\beta}^2} \tag{7.3.15}$$

and $$\hat{\alpha} = \hat{\lambda}\log(1-\hat{\beta}) - \log(n_1/n_0) + \log\hat{\theta} \tag{7.3.16}$$

(c) Fitting of LCPD-I

Some reported observed data have been considered for the fitting of the four parameter LCP distribution of type-I for empirical comparison. In

Table7.2, the accidents data of shunting service for different age group have been considered for which Adelstein (1952) had used the Poisson and negative binomial distribution and Consul (1989) had successfully fitted the GPD model. It may be noted that when LCPD-I model has been applied to these data, it provides an excellent fit to all the sets of data as judged by the chi-square values in Table7.2.

In Table 7.3, we consider Kendall (1961) data for fitting of LCPD-I model. Kendall (1961) considered the observed data on the number of strikes in 4-week period in four leading industries in U.K. during 1948-1959 and concluded that the aggregate data for the four industries agrees with Poisson law but it did not hold well for the individual industries. The LCPD-I has been fitted to the observed data for the four individual industries and the results are given in Table 7.3 along with the expected frequencies. Based on the expected frequencies and the corresponding χ^2 values (Table 7.3), it is clear that the pattern of strikes in coal mining, vehicle manufacture, ship building and transport industries follow LCPD model and this distribution gives better fit than the GPD model for coal mining industries. In the light of the above discussions it may be stated that the Charlier type of Lagrangian probability distribution can be applied in various fields of experiment with varied amount of success.

Table 7.2 Comparison of Observed Frequencies for First- Year Shunting Accidents and for a Five-Year Record of Experienced Men with Expected LCPD-I and GPD Frequencies for Different Age Group.

Number of Accidents	Age 21-25 yr.			Age 26-30 yr.			Age 31-36			5- yr. Record for experienced		
	Observed	LCPD	GPD	Observed	LCPD	GPD	Observed	LCPD	GPD	Observed	LCPD	GPD
0	80	79.99	76.40	121	120.99	126.42	80	79.99	80.23	54	54.00	51.77
1	56	55.99	65.03	85	84.99	74.49	61	60.99	60.41	60	59.99	62.11
2	30	29.84	23.37	19	18.66	21.45	13	12.99	13.48	36	38.54	40.02
3	4	4.05	5.2	1	1.89	4.02	1	1.02	0.88	21	18.69	18.41
≥ 4	0	.08		1	0.31	0.62	0			11	10.55	9.69
Total	170			227			155			182		
		$\hat{\theta}$ =0.75			0.63			0.66			1.21	
and		$\hat{\alpha}$ =-0.18			-0.14			-0.11			0.10	
Estimates		$\hat{\beta}$ =-1.09			0.46			-0 26			0.76	
		$\hat{\lambda}$ =-0.34			0.05			0.12			-.008	
		χ^2 = .0049	3.58		0.024	3.45		0.0008	0.04		0.46	1.13

Table 7.3 Comparison of Observed Frequencies of the Number of Outbreaks of Strike in Four Leading Industries in U.K. During 1948-1959 with the Expected LCPD-I and GPD Frequencies.

Number of outbreaks	Coal mining			Vehicle manufacturing			Ship building			Transport		
	Observed	LCPD	GPD	Observed	LCPD	GPD	Observed	LCPD	GPD	Observed	LCPD	GPD
0	46	45.99	50.01	110	110.00	109.82	117	117.00	116.74	114	114.00	114.84
1	76	75.99	66.77	33	32.94	33.36	29	28.99	30.22	35	34.99	33.88
2	24	22.55	31.23	9	9.30	9.24	9	9.39	6.97	4	3.34	7.27
3	9	11.20	7.23	3	2.65	3.58	0	2.03	0.88	2	2.35	2.01
4	1	1.15	0.76	1	0.76		1	0.81		1	1.64	9.69
Total	156			156			156			156		
and		$\hat{\theta}$ =1.22			0.34			0.28			0.31	
Estimates		$\hat{\alpha}$ =0.24			0.15			0.09			0.02	
		$\hat{\beta}$ =-0.34			0.82			-3.82			-0.75	
		$\hat{\lambda}$ =1.85			-0.001			-0.04			0.42	
		χ^2 =0.29	χ^2 =4.52		1.47	0.06		1.21	1.19		0.25	2.27

References

Adelstein, A.M. (1952). Accident proneness: a criticism of the concept based upon an analysis of shunters' accidents, *Journal of the Royal Statistical Society, Series A*, **115**, 354-410.

Anruku, K. and Yanagimoto, T. (1990). Estimation for the negative binomial distribution based on the conditional likelihood, *Communications in Statistics, Simulation and Computation*, **19**,771-806.

Beall, G. (1940). The fit and significance of contagious distributions when applied to observations on larval insects, *Ecology*, **21**, 460-474.

Beall, G. and Rescia, R.R. (1953). A generalization of Neyman's contagious distributions, *Biometrics*, **9**, 354-386.

Berg, S.(1985). Generating discrete distributions from modified Charlier type B expansions, *Contribution to Probability and Statistics in Honour of Gunnar Blom*, J.Lanke and G. Lindgren (editors), 39-48, Lund: University of Lund.

Bernoulli, J. (1713). *Ars Conjectandi*, Basilea: Thurnisius.

Bhattacharya, S.K. (1966). Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness, *Bulletin of the Calcutta Statistical Association*, **15**, 20-31.

Bliss, C.I., and Fisher, R.A.(1953). Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial, *Biometrics*, **9**, 176-200.

Borah, M.(1979). Some properties of moments and cumulants of Bionomial-Poisson distribution and its limiting distributions, *Pure and applied Mathematika Sciences*, **9**, 65-69.

Borah, M.(1980). On certain recurrence relation arising in Poisson-Binomial distributions, *Journal of Assam Science Society*, **23**, 50-53.

Borah, M.(1984). The Gegenbauer distribiution revisited: Some recurrence relationsfor moments, cumulants, etc. estimation of parameters and its goodness of fit, *Journal of Indian Society of Agricultural Statistics*, **36**, 72-78.

Borah, M. and Begum, R.A.(1997). A class of Discrete Probability distributions applicable in Queueing Theory, *proceeding of Mathematics and Statistics in Engineering and Technology*, A. Chattopadhyay (editor).

Borah, M. and Deka Nath, A. (2000). A Class of Hermite Type Lagrangian Distribution, *Recent Trends in Mathematical Sciences*, J.C. Misra and S.B.Sinha (editors), Narosa Publishing house, 319-326.

Borah, M. and Deka Nath, A. (2001). Poisson-Lindley and some of its Mixture distributions, *Journal of Pure and Applied Mathematika Sciences*, Vol. **53**, No. 1-2, 1-8.

Borah, M. and Deka Nath, A. (2001). A Study on the Inflated Poisson Lindley Distribution, *Journal of Indian Society of Agricultural Statistics,* **54(3)**, 317-323.

Bowman, K.O. Kastenbaun, M.A., and Shenton, L.R. (1992). The negative hypergeometric distribution and estimation by moments, *Communications in Statistics, Simulation and Computation,* **21**, 301-332.

Chang, D.K. (1989). On infinitely divisible discrete distributions, *Utilitas Mathematica*, **36**, 215-217.

Charalambides, C.A. (1986). Derivation of probabilities and moments of certain generalized discrete distributions via urn models, *Communications in Statistics- Theory and Methods*,**15**, 677-696.

Chatfield, C. (1969). On estimating the parameters of the logarithmic series and negative binomial distributions, *Biometrika*, **56**, 411-414.

Clark, S.J. and Perry, J.N. (1989). Estimation of the negative binomial parameter k by maximum quasi-likelihood, *Biometrics*, **45**, 309-316.

Cohen, A.C. (1960). Estimating the parameters of a modified Poisson distribution, *Journal of the American Statistical Association*, **55**, 139-143.

Cohen, A.C.(1963). An extension of a truncated Poisson Distribution, *Biometrics*,**16,** 446-450.

Consul, P.C. (1975). Some new characterizations of discrete Lagrangian distributions, *Statistical Distributions in Scientific Work,* **3**: Characterizations and Applications, G.P. Patil, S. Kotz and J.K. Ord (editors), 279-290.

Consul, P.C. (1981). Relation of modified power series distributions in Lagrangian probability distributions, *Communications in Statistics-Theory and Methods, A,* **10**, 2039-2046.

Consul, P.C. (1983). Lagrange and related probability distributions, *Encyclopedia of Statistical Sciences*, **4**, S. Kotz, N.L. Johnson and C.B. Read (editors), 448-454 New York: Wiley.

Consul, P.C. (1989). *Generalized Poisson Distributions Properties and Applications*, New York, Dekker.

Consul, P.C. (1990). Geeta distribution and its properties, *Communications in Statistics-Theory and Methods*, **19**, 3051-3068.

Consul, P.C. and Famoye, F. (1989). Minimum variance unbiased estimation for the Lagrange power series distributions, *Statistics*, **20**, 407-415.

Consul, P.C. and Felix, F. (1996). Lagrangian katz family of distributions, *Communications in Statistics-Theory and Methods,* **25(2),** 415-434.

Consul, P.C. and Gupta, H.C. (1980). The generalized negative binomial distribution and its characterization by zero regression, *SIAM Journal of Applied Mathematics*, **39**, 231-237.

Consul, P.C. and Jain, G.C. (1973). A generalized Poisson distribution, *Technometrics*, **15**, 791-799.

Consul, P.C. and Shenton, L.R. (1972). Use of Lagrange expansion for generating generalized probability distributions, *SIAM Journal of Applied Mathematics*, **23**, 239-248.

Consul, P.C. and Shenton, L.R. (1973). Some interesting properties of Lagrangian distributions, *Communications in Statistics*, **2**, 263-272.

Consul, P.C. and Shenton, L.R. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions, *Statistical Distributions in Scientific Work*, **1**: Models and Structures, G.P. Patil, S. Kotz, and J. K. Ord (editors), 41-48, Dordrecht: Reidel.

Cresswell, W.L. and Froggatt, P. (1963). The Causation of Bus Driver Accidents, London: Oxford University Press.

Dawid, A.P. and Skene, A.M. (1979). Maximum likelihood Estimation of Observer Error-Rates using the EM algorithm, *Applied Statistics*, **28**, 20-28.

Deka Nath, A. and Borah, M. (2000). The Short Poisson-Poisson-Lindley Distribution, *Journal of Assam Science Society*, Vol **41**, No. 2, 120-128.

Deka Nath, A. and Borah, M. (2001). A study on Charlier Type Lagrangian probability distributions, *Proceedings of the National Conference on Mathematical and Computational Models*, December 27-28, PSG College of Technology, Coimbatore, India, 67-72.

Doetsch, G. (1934). Die in der Statistik seltener Ereignisse auftrewtenden Charlierschen Polynome und eine damit zusammehhangenda Differential-differenzengleichung. *Math. Ann.* 109, 257-266.

Douglas, J.B. (1955). Fitting the Neyman type A (two parameter) contagious distribution, *Biometrics*, **11**, 149-173.

Douglas, J.B. (1965).Asymptotic expressions for some contagious distributions, *Classical and Contagious Discrete Distributions*, G.P, Patil (editor), 291-302. Calcutta Statistical Publishing Society; Oxford: Pergamon Press.

Douglas, J.B. (1980). *Analysis with Standard Contagious Distributions*, Burtonsville, MD: International Co-operative Publishing House.

Evans, D.A. (1953). Experimental evidence concerning contagious distributions in ecology, *Biometrika*, **40**, 186-211.

Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*, London: Chapman and Hall.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 1 (Third edition), New York: Wiley.

Gerstenkorn, T. (1979). Moment recurrence relations for the inflated negative binomial, Poisson and geometric distributions, *Demonstratio Mathematica*, **11**, (2), 389-411.

Gokhale, D.V. (1980). On power series distributions with mean- variance relationships, *Journal of the Indian Statistical Association*, **18**, 81-84.

Greenwood, M. and Yule, G.U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal Statistical Society, Series A*, **83**, 255-279.

Grzegorska, L. (1973). Recurrence relations for the Moments of the so-called Inflated Distributions, *Annales Universitatis Mariae Curie-Sklodowska, Lublin*, Vol. **XXVII**, 3, 19-29.

Gupta, R.C. (1970). Some Characterizations of Probability Distributions. Ph.D thesis, Detroit, *MI: Wayne State University*.

Gupta, R.C. (1974). Modified power series distributions and some of its applications, *Sankhya, Series B*, **35**, 288-298.

Gupta, R.C. (1975). Maximum-likelihood estimation of a modified power series distribution and some of its applications, *Communications in Statistics*, **4**, 689-697.

Gupta, R.C. (1977). Minimum variance unbiased estimation in modified power series distribution and some of its applications, *Communications in Statistics-Theory and Methods*, **A 6**, 977-991.

Gupta, R.C. and Singh, J. (1982). Estimation of probabilities in the class of modified power series distributions, *Mathematicsche Operationforschung und Statistik, Series Statistics*, **13**, 71-77.

Gupta, R.C. and Tripathi, R.C. (1985). Modified power series distributions, *Encyclopedia of Statistical Sciences*, **5**, S. Kotz, N.L. Johnson and C.B. Read (editors) 593-599, New York: Wiley.

Gupta, R.C. and Jain, G.C. (1974). A generalized Hermite distribution and its properties, *SIAM Journal of Applied Mathematics*, **27**, 359-363.

Gurland, J. (1957). Some interrelations among compound and generalized distributions, *Biometrika*, **44**, 265-268.

Gurland, J. (1958). A generalized class of contagious distributions, *Biometrics*, **14**, 229-249.

Gurland, J. (1965). A method of estimation for some generalized Poisson distributions, *Classical and Contagious Distributions*, G.P. Patil (editor), 141-158, Calcutta: Statistical Publishing Society; Oxford: Pergamon Press.

Haight, F.A. (1961). A distribution analogous to the Borel-Tanner, *Biometrika*, **48**, 167-173.

Harris, J.F. (1983). On finite mixtures of geometric and negative binomial distributions, *Communications in Statistics-Theory and Methods*, **12**, 987-1007.

Holgate, P. (1964). A modified geometric distribution arising in trapping studies, *Acta Theriologica*, **9**, 353-356.

Ishii, G., and Hayakawa, R.(1960). On the compound binomial distribution, *Annals of the Institute of Statistical Mathematics*, Tokyo, **12**, 69-80.

Jain, G.C.(1975). A power series distribution association with Lagrange expansion, *Biometrische Zeitsehrift*, **17**, 85-97.

Jain, G.C. and Consul, P.C. (1971): A generalized negative binomial distribution, *SIAM Journal of Applied Mathematics*, **21**, 501-513.

Jain, G.C. and Plunkett, I.G. (1977). A Borel-Hermite distribution and its applications, *Biometrical Journal*, **19**, 347-354.

Janardan, K.G. (1978). On generalized Markov-Polya distribution, *Gujurat Statistical Review*, **5**, 16-32.

Janardan, K.G. (1998). Generalized Polya Eggenberger family of distributions and its relation to Lagrangian Katz family, *Communications in Statistics-Theory and Methods*, **27(10)**, 2423-2443.

Janardan, K.G. and Rao, B.R. (1982). Characterization of generalized Markov-Polya and generalized Polya-Eggenberger distributions, *Communications in Statistics*, **11**, 2113-2124.

Janardan, K.G. and Rao, B.R (1983). Lagrangian distribution of the second kind and weighted distributions, *SIAM Journal of Applied Mathematics*, **43**, 2, 302-313.

Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions* (First edition), Boston: Houghton Mifflin.

Johnson, N.L. Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Doistributions*, (second edition), John Wiley & Sons, New York.

Katti, S.K. and Gurland, J. (1961). The Poisson Pascal distribution, *Biometrics*, **17**, 527-538.

Katti, S.K. and Gurland, J. (1962a). Efficiency of certain methods of estimation for the negative binomial and the Neyman type A distributions, *Biometrika,* **49**, 215-226.

Katti, S.K. and Gurland, J. (1962b). Some methods of estimation for the Poisson binomial distribution, *Biometrics*, **18**, 42-54.

Katti, S.K. and Rao, A.V. (1970). The log-zero-Poisson distribution, *Biomatrics*, **26**, 801-813.

Kemp, A.W. (1986). Weighted discrepancies and maximum likelihood estimation for discrete distributions, *Communications in Statistics Theory and Method,* **15**, 783-803.

Kemp, A.W. and Kemp, C.D. (1966). An alternative derivation of the Hermite distribution, *Biometrika*, **53**, 627-628.

Kemp, C.D. (1967). On contagious distribution suggested for accident data, *Biometrics*, **23**, 241-255.

Kemp, C.D. and Kemp, A.W. (1965). Some properties of the 'Hermite' distribution, *Biometrika*, **52**, 381-394.

Kemp, C.D. and Kemp, A.W. (1988). Rapid estimation for discrete distributions, *The Statistics*, **37**, 243-255.

Kendall, M.G. (1961). Natural law in the social sciences, *Journal of the Royal Statistical Society, Series A*, **124**, 1-18.

Khatri, G.C. (1959). On certain properties of power-series distributions, *Biometrika*, **46**, 486-490.

Khatri, G.C.(1961). On the distributions obtained by varying the number of trials in a binomial distribution, *Annals of the Institute of Statistical Mathematics*, Tokyo, **13**, 47-51.

Khatri, G.C. and Patel, I.R. (1961). Three classes of univariate discrete distributions, *Biometrics*, **17**, 567-575.

Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, **20**, 102-107.

Lingappaiab, G.S. (1977). On the Inflated Discrete Distribution, *Statistica (Bologna)*, Vol. **37 (3)**, 379-386.

Lundberg, O. (1940). *On Random Process and their Applications to Sickness and Accident Statistics*, Uppsala: Almqvist and Wicksells (Repriented 1964).

Maceda, E.C. (1948). On the compound and generalized Poisson distributions, *Annals of Mathmatical Statistics*, **19**, 414-416.

Martin, D.C. and Katti, S.K. (1965). Fitting of some contagious distributions to some available data by the maximum likelihood method, *Biometrics*, **21**, 34-48 (correction, 21, 514).

McGuire, J.U. Brindley, T.A. and Bancroft, T.A. (1957). The distribution of European corn borer Pyrausta Nubilalis (Hbn) in field corn, *Biometrics*, **13**, 65-78 (extra and extensions 14 (1958), 432-434.).

Medhi, J. and Borah, M. (1984). On generalized Gegenbauer polynomials and associated probabilities, *Sankhyā*, *Series B*, **46**, 157-165.

Medhi, J. and Borah, M. (1986). On generalized four-parameter Charlier distribution, *Journal of Statistical Planning and Inferencas*, North-Holland, **14**, 69-77.

Meixner, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion, *Journal of London Mathematical Society*, **9**, 6-13.

Meixner, J. (1938). Erzeugende Funktionen der Charlierschen Pelynome, *Math. Z.*, **44**, 531-535.

Mohanty, S.G. (1966). On a generalized Two-coin tossing problem, *Biometrische Zeitschrift*, **8**, 266-272.

Neyman, J. (1939). On a new class of "contagious" distributions applicable in entomology and bacteriology, *Annals of Mathematical Statistics*, **10**, 35-57.

Noack, A. (1950). A Class of Random Variables with Discrete Distributions, *Annals of Mathematical Statistics*, **21**, 127-132.

Pandey, K.N. (1965). On generalized inflated Poisson distribution, *Banaras Hindu University Journal of Scientific Research*, **15**, 157-162.

Patel, I.D. (1975). A note on an Inflated Power Series distribution, *Acta Ciencia India*, **1**, No. 2, 133-135.

Patel, Y.C. (1971). *Some problems in Estimation for the Parameters of the Hermite Distribution*, Ph.D. dissertation, Athens, GA; *University of Georgia*.

Patel, Y.C. (1976). Even point estimation and moment estimation in Hermite distribution, *Biometrics*, **32**, 865-873.

Patel, Y.C. Shenton, L.R. and Bowman, K.O. (1974). Maximum-likelihood estimation for the parameters of the Hermite distribution, *Sankhyā*, *Series B*, **36**, 154-162.

Patil, G.P. (1961). *Contributions to Estimation in a Class of Discrete Distributions*, Ph.D. thesis, Ann Arbor, MI: University of Michigan.

Patil, G.P. (1962a). Maximum-likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution, *Biometrika*, **49**, 227-237.

Patil, G.P. (1962b). Some method of estimation for the logarithmic series distribution, *Biometrices*, **18**, 68-75.

Patil, G.P. (1964). On certain compound Poisson and compound binomial distributions, *Sankhyā*, *Series A*, **26**, 293-294.

Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size based sampling with applications to wildlife populations and human families, *Biometrics*, **34**, 179-189.

Patil, S.A. and Raghunandanan, K. (1990). Compound Hermite and stuttering Poisson distributions, *Bulletin of the Calcutta Statistical Association*, **39**, 97-103.

Pearson, K. (1915). On certain types of compound frequency distributions in which the components can be individually described by binomial series, *Biometrika*, **11**, 139-144.

Plunkett, I.G. and Jain, G.C. (1975). Three generalized negative binomial distributions, *Biometriche Zeitschrift*, **17**, 276-302.

Poisson, S.D. (1837). *Recherches sur la Probabilite des Jugements en Matiere Criminelle et en Mattere Civile, Precedees des Regles Generales du Calcul des Probabilities*. Paris; Bachelier, Imprimeur-Libraire pour les Mathematiques, la Physique etc.

Polya, G. (1930). Sur quelques points de la theorie des probabilities, *Annales de l'Institut H. Poincare*, **1**, 117-161.

Russell, K.G. (1978). Estimation of parameters of the Thomas distribution, *Biometrics*, **34**, 369-411.

Sankaran, M. (1970). The discrete Poisson-Lindley distribution, *Biometrics*, **26**, 145-149.

Shumway, R., and Gurland, J. (1960). Fitting of Poisson binomial distribution, *Biometrics*, **16**, 522-533.

Singh, M.P. (1965). Inflated binomial distribution, *Banaras Hindu University Journal of Scientific Research*, **16**, (1), 87-90.

Singh, M.P. (1966): A note on generalized inflated binomial distribution. *Sankhya Ser.* A., **28**, (1),99

Singh, S.N. (1963). A note on inflated Poisson distribution, *Journal of the Indian Statistical Association*, **1**, 140-144.

Skellam, J.G. (1952).Studies in statistical ecology 1: Spatial pattern, *Biometrika*, **39**, 346-362.

Smith, A.F.M. (1985). An overview of some problems relating to finite mixtures, *Rassegna di Metodi Statistici ed Applicatione 5 Cagliari*, 138-149.

Sobic, L. and Szynal, D. (1974). Some properties of Inflated Binomial distribution, *Canadian Mathematical Bullitin*, Vol. **17 (4)**, 611-613.

Titterington, D.M. (1990). Some recent research in the analysis of mixture distributions, *Statistics*, **21**, 619-641.

Tripathi, R.C. Gupta, P.L. and Gupta, R.C. (1986). Incomplete moments of modified power series distributions with applications, *Communications in Statistics-Theory and Methods*, **15**, 999-1015.

Tripathi, R.C. and Gupta, R.C. (1985). A generalization of the log-series distribution, *Communications in Statistics-Theory and Methods*, **14**, 1779-1799.

Tripathi, R.C. and Gupta, R.C. (1988). Another generalization of the logarithmic series and the geometric distributions, *Communications in Statistics-Theory and Methods*, **17**, 1541-1547.

Tripathi, R.C. Gurland, J., and Bhalerao, N.R. (1986). A unified approach to estimating parameters in some generalized Poisson distributions, *Communications in Statistics-Theory and Methods*, **15**, 1017-1034.

Williams, C.B. (1944). The number of publications written by biologists, *Annals of Eugenics*, **12**, 143-146.

Williams, C.B. (1947). The logarithmic series and its application to biological problems, *Journal of Ecology*, **34**, 253-272.

APPENDIX A

The following papers based on the work done in this thesis have been published / accepted for publication / communicated for publication in different journals

Published Papers:

1. Borah, M. and Deka Nath, A.(2000). A class of Hermite type Lagrangian Distributions, Recent Trends in Mathematical Sciences, J.c. Misra and S.B. Sinha (editors), Narosa Publishing House, 319-326.

2. Borah, M. and Deka Nath, A.(2001). A study on the Inflated Poisson Lindley Distribution, Journal of the Indian Society of Agricultural Statistics, 54 (3), 317-323.

3. Borah, M. and Deka Nath, A.(2001). Poisson Lindley and Some of its Mixture distributions, Pure and applied Mathematika Sciences, 53, No. 1-2,1-8.

4. Deka Nath, A. and Borah, M. (2000). The Short Poisson-Poisson-Lindley Distribution, Journal of Assam Science Society, 41, No.2, 120-128.

5. Deka Nath, A. and Borah, M. (2001). A study on Charlier Type Lagrangian Probability Distribution, Proceeding of National Conference on Mathematical and Computational Models, December 27-28, PSG College of Technology, Coimbatore, INDIA, 67-72.

Accepted for Publication:

1. Deka Nath, A. and Borah, M.; A Study on Inflated Geometric Distribution, Proceeding of 37th Annual Technical Sessions of Assam Science Society.

2. Deka Nath, A. and Borah, M.; On certain Recurrence Relations Arise in some Discrete Inflated Probability Distributions, Journal of Assam Statistical Review.

Communicated for Publications:

1. Deka Nath, A.; A study on Inflated Power series Distribution, Sankhya Series A.

APPENDIX B

- COPIES OF PUBLISHED PAPERS OF THE AUTHOR

34. A Class of Hermite Type Lagrangian Distributions

M. Borah
Dept. of Mathematical Sciences
Tezpur University, Napaam
Tezpur-784025, Assam (India)

A. Deka Nath
Department of Statistics
Darrang College,
Tezpur-784001, Assam (India)

Abstract

A class of Hermite type of Lagrangian probability distributions have been defined by using well known Lagrange's expansions. The probability mass function and cumulants of the basic Lagrangian Hermite (LHD) distribution are provided. The parameters are estimated by using method of moments and method of first frequency and mean. Some applications of this distribution are also considered. Then several members of Lagrangian Hermite type distributions of type I and type II are investigated by various choice of probability generating function of f (t) and g (t). For example, the Lagrangian Hermite Poisson distribution of type I and type II are derived and fitted to some well-known data with good results.

Key Words: Lagrange's Expansion, Lagrangian Probability Distribution, Hermite distribution and Cumulants.

1. Introduction

Lagrangian expansion for the derivation of the probabilities of certain discrete distributions has been used by Consul and Shenton [1], [2], [3] and Mohanty [4], Consul and Jain [5] and their co-workers. The nature of the generalization process for these distributions were clarified by Consul and Shenton [1], [2] and also by Consul [6] in their papers.

If g(t) and f(t) are two probability generating functions (pgf) defined on non negative integers such that $g(0) \neq 0$, then the pgf for the general Lagrangian distribution formed from g(t) and f(t) by considering the transformation $t = u\,g(t)$, is given by

$$f(t) = f(0) + \sum_{s=1}^{\infty} \frac{u^s}{s!} \frac{d^{s-1}}{dt^{s-1}} \left\{ (g(t))^s \, f'(t) \right\} \Big|_{t=0} \qquad (1.1)$$

Thus the probability mass function (pmf) for the Lagrangian probability distribution is given by

$$P_r[X=x]=\frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\left\{(g(t))^x . f'(t)\right\}\Big|_{t=0} \qquad x=1,2,3,.... \qquad (1.2)$$

where $P_r[X=0]=f(0)$

Equation(1.2) is also known as Lagrangian probability distribution of type-I(LPD-I) (according to Janardan and Rao's terminology).

Using Lagrangian expansion of 2nd kind Janardan and Rao [7] investigated a new class of discrete distribution called the Lagrangian probability distribution of type -II (LPD-II), with pmf

$$P_r(X=x)=\frac{\{1-g'(1)\}}{x!}\left[\frac{\delta^x}{\delta z^x}\left\{(g(z))^x f(z)\right\}\right]_{z=0}, \quad for\ x=0,1,2,... \qquad (1.3)$$

$= 0$, otherwise

The Hermite distribution was formally introduced by Kemp and Kemp [8] and was applied in the field of biological sciences, physical sciences and operation research. Hermite distribution is a generalized Poisson distribution whose pgf is

$$\exp\left[\alpha_1(t-1)+\alpha_2(t^2-1)\right]$$

The probabilities of which can be conveniently expressed in terms of modified Hermite polynomials.

The motivation behind this paper is to derive the basic Lagrangian Hermite distribution (LHD). The cumulants of this distribution are investigated. The parameters of this distribution are estimated by method of moments and ratio of first frequency and mean. Considering different values of f(t) and g(t) in equation (1.2) and (1.3), different general Lagrangian Hermite type distributions are generated and Lagrangian Hermite Poisson distribution of type-I and type-II are particularly investigated in this paper.

2(a). Basic Lagrangian Hermite distribution (LHD)

The pmf of basic Lagrangian distribution is given as

$$P_r[X=x]=\frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\{g(t)\}^x\Big|_{t=0} \qquad x=1,2,3,... \qquad (2.1)$$

$= 0$, otherwise.

where g(t) is the pgf defined on some or all non negative integers, such that $g(0)\neq 0$. In this case

$$g(t)=\exp\left[\alpha_1(t-1)+\alpha_2(t^2-1)\right] \qquad (2.2)$$

Thus the pmf of basic LHD may be written as

$$P_r[X=x]=\exp[-x(\alpha_1+\alpha_2)]\sum_{j=0}^{[(x-1)/2]}\frac{\alpha_1^{x-1-2j}\alpha_2^{j}x^{x-1-j}}{(x-2j)!\,j!}, \qquad x=1,2,3\ldots \qquad (2.3)$$

= 0, otherwise.

where $[(x-1)/2]$ denotes the integer part of (x-1)/2

(b) Cumulants of the basic LHD

The cumulants of basic Lagrangian distribution are investigated by using Consul and Shenton [3] general formula. The moments can be directly obtained by cumulants. If G_i be the ith cumulants of the distribution with pgf g(t) then the first four cumulants of basic Lagrangian distribution can be written as

$$k_1=\frac{1}{1-G_1}$$

$$k_2=\frac{G_2}{(1-G_1)^3}$$

$$k_3=\frac{G_3}{(1-G_1)^4}+\frac{3G_2^{\,2}}{(1-G_1)^5}$$

$$k_4=\frac{G_4}{(1-G_1)^5}+\frac{10G_3G_2}{(1-G_1)^6}+\frac{15G_3^{\,2}}{(1-G_1)^7}$$

In case of Hermite distribution we have the cumulants as

$$G_1=\alpha_1+2\alpha_2 \qquad (2.4)$$

$$G_2=\alpha_1+4\alpha_2 \qquad (2.5)$$

$$G_3=\alpha_1+8\alpha_2 \qquad (2.6)$$

$$G_4=\alpha_1+16\alpha_2 \qquad (2.7)$$

Thus the first four cumulants of basic LHD may be written as

$$k_1=\frac{1}{(1-\alpha_1-2\alpha_2)} \qquad (2.8)$$

$$k_2=\frac{\alpha_1+4\alpha_2}{(1-\alpha_1-2\alpha_2)^3} \qquad (2.9)$$

$$k_3=\frac{\alpha_1+8\alpha_2}{(1-\alpha_1-2\alpha_2)^4}+\frac{3(\alpha_1+4\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^5} \qquad (2.10)$$

$$k_4 = \frac{\alpha_1 + 16\alpha_2}{(1-\alpha_1-2\alpha_2)^5} + \frac{10(\alpha_1+8\alpha_2)(\alpha_1+4\alpha_2)}{(1-\alpha_1-2\alpha)^6} + \frac{15(\alpha_1+8\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^7} \quad (2.11)$$

(c) Estimation of parameter

Method of moments and ratio of first frequency and mean can be used to estima the parameters of basic LHD.

(i)Method of moments:

The mean and variance of the basic LHD, as given in (2.8) and (2.9) may be writt as

$$\bar{x} = \frac{1}{(1-\alpha_1-2\alpha_2)} \quad (2.12)$$

$$m_2 = \frac{\alpha_1 + 4\alpha_2}{(1-\alpha_1-2\alpha_2)^3} \quad (2.13)$$

By eliminating α_1 between (2.12) and (2.13) we may obtain

$$\hat{\alpha}_2 = \frac{1}{2}\left[\frac{m_2}{\bar{x}^3} + \frac{1}{\bar{x}} - 1\right] \quad (2.14)$$

Substituting the value of $\hat{\alpha}_2$ in equation (2.12) we may get

$$\hat{\alpha}_1 = 2 - \frac{m_2}{\bar{x}^3} - \frac{2}{\bar{x}} \quad (2.15)$$

(ii) Ratio of first frequency and mean:

By equating the first probability of basic LHD with $\frac{n_1}{N}$ we may obtain

$$\frac{n_1}{N} = \exp(-\alpha_1 - \alpha_2) \quad (2.16)$$

By eliminating α_1 between (2.12) and (2.16) we may obtain

$$\hat{\alpha}_2 = 1 - \frac{1}{\bar{x}} + \log\frac{n_1}{N} \quad (2.17)$$

hence, $$\hat{\alpha}_1 = \frac{1}{\bar{x}} - 2\log\frac{n_1}{N} - 1 \quad (2.18)$$

(d) Fitting of basic LHD

For the application of basic LHD, we consider the example of number of pap published per author in the review of applied entomology data by Kendall [9] for wh geometric distribution (GD), logarithmic series distribution (LSD) are fitted by Williams [10] and generalized logarithmic series distribution (GLSD) by Jain [11]. The observed and the expected frequencies of this example are given in Table 1. It is clear from Table 1 that the LHD model gives much better fit than GD, LSD and GLSD as shown by the values of χ^2.

Table 1. Fitting of no. of papers per author by LHD, GLSD, LSD and GD. Publication in the review of applied entomology. Vol. 24, 1936 (2379 papers by 1534 authors)

No. of papers per author	**Observed** frequency	**LHD** $\alpha_1 = 0.3803$ $\alpha_2 = -0.012$	**GLSD** Jain (1975)	**LSD** Williams (1944)	**GD** Williams (1944)
1	1062	1062.06	1052.72	1046.05	989.10
2	263	279.59	287.52	293.05	351.30
3	120	104.02	107.10	109.46	124.80
4	50	45.95	45.10	45.99	44.33
5	22	21.33	20.83	20.61	15.75
6	7	10.33	10.00	9.62	5.59
7	6	5.45	4.97	4.62	1.99
8	2	2.30	2.53	2.26	0.71
9	0	1.23	1.31	1.12	0.25
10	1	1.19	0.70	0.53	0.09
11	1	0.55	0.81	0.66	0.09
Total	1534	1534.00	1534.00	1534.00	1534.00
χ^2		4.94	5.14	5.56	46.39

3(a). General Lagrangian Hermite Poisson Distribution:

Considering different values of g(t) and f(t) in (1.2) and (1.3) different LHD of type I and type II may be obtained. Let

$$g(t) = \exp[\alpha_1(t-1) + \alpha_2(t^2-1)]$$

$$f(t) = \exp[\theta(t-1)]$$

Hence the pmf of Lagrangian Hermite Poisson distribution of type I (LHPD-I) may be written as

$$P_r(X = x) = \theta \exp\{\theta + x(\alpha_1 + \alpha_2)\} \sum_{j=0}^{\left[\frac{x-1}{2}\right]} \frac{(\theta + x\alpha_1)^{x-1-2j} \alpha_2^j x^j}{(x-2j)!\, j!} \quad x=1,2,3,\ldots \qquad (3.1)$$

$$P_r(X=0)=e^{-\theta}$$

where $\theta > 0$.

Similarly considering (1.3) the p.m.f. of LHD-II may be written as

$$P_r(X=x)=A\exp\{\theta+x(\alpha_1+\alpha_2)\}\sum_{j=0}^{[x/2]}\frac{(\theta+x\alpha_1)^{x-2j}\alpha_2^{\,j}x^j}{(x-2j)!\,j!} \qquad x=0,1,2,\ldots \qquad (3.2)$$

and zero otherwise, where A={1-($\alpha_1+2\alpha_2$)}

(b) Cumulants of general LHPD-I

According to Consul and Shenton [3] if F_r be the rth cumulants for the pgf f(t) as a function of z, and if D_r be the rth cumulants for the basic Lagrangian distribution obtained from g(t) then the cumulants of general Lagrangian distribution may be written as

$$k_1=F_1D_1$$
$$k_2=F_1D_2+F_2D_1^2$$
$$k_3=F_1D_3+3F_2D_1D_2+F_3D_1^3$$
$$k_4=F_1D_4+3F_2D_1^2+4F_2D_1D_3+6F_3D_1^2D_2+F_4D_1^4$$ (See Consul and Shenton [3])

Here D_1, D_2, D_3 and D_4 are given in equation (2.8), (2.9), (2.10) and (2.11) respectively. Thus

$$k_1=\frac{\theta}{(1-\alpha_1-2\alpha_2)} \qquad (3.3)$$

$$k_2=\frac{\theta(\alpha_1+4\alpha_2)}{(1-\alpha_1-2\alpha_2)^3}+\frac{\theta}{(1-\alpha-2\alpha_2)^2} \qquad (3.4)$$

$$k_3=\frac{3\theta(\alpha_1+4\alpha_2)^2}{(1-\alpha_1-2\alpha_2)^5}+\frac{4\theta(\alpha_1+5\alpha_2)}{(1-\alpha_1-2\alpha_2)^4}+\frac{\theta}{(1-\alpha_1-2\alpha_2)^3} \qquad (3.5)$$

Hence the parameters of LHPD-I can be estimated in terms of its cumulants.

(c) Estimation of parameters

The parameters of LHPD-I may be estimated from the ratio of first two frequencies and the mean. By equating the first and second probabilities of LHPD-I with n_0/N and n_1/N, we may obtain

$$\hat{\theta}=-\log\frac{n_0}{N}, \qquad (3.6)$$

and $$\log\frac{n_1}{N} = \log\theta - (\theta + \alpha_1 + \alpha_2), \tag{3.7}$$

from (3.3) we have $$\bar{x} = \frac{\theta}{1 - \alpha_1 - 2\alpha_2} \tag{3.8}$$

Hence from equation (3.6), (3.7) and (3.8) we may obtain the estimate of α_1 and α_2 as

$$\hat{\alpha}_2 = 1 - \frac{\hat{\theta}}{\bar{x}} + \log\frac{n_1}{N} - \log\hat{\theta} + \hat{\theta} \tag{3.9}$$

and $$\hat{\alpha}_1 = 1 - \frac{\hat{\theta}}{\bar{x}} - 2\hat{\alpha}_2 \tag{3.10}$$

(c) Fitting of LHPD-I

In Table 2 we consider Adelstein [12], data on number of accidents (home injuries) of 122 experienced men in 11 years period where Adelstein had concluded that the Poisson distribution fits the first sets of data but did not fit the second and third sets. When the LHPD-I distribution is applied to those sets it provides good fit in all the cases. In Table 2 we get $\theta^* = 0.5135$, $\alpha_1^* = 0.0103$ and $\alpha_2^* = 0.0304$ for the 1st set and $\theta^* = 0.3267$, $\alpha_1^* = -0.1799$ and $\alpha_2^* = 0.648$ for the 2nd set and $\theta^* = 0.7435$, $\alpha_1^* = 0.006$ and $\alpha_2^* = 0.2316$ for 3rd set.

Table 2. Home injuries of 122 Experienced Men during 11 years with expected LHPD-I Frequencies [Adelstein [12] data].

No. of	1937-1942		1943-1947		1937-1947	
Injuries	Observed	LHPD-I	Observed	LHPD-I	Observed	LHPD-I
0	73	72.99	88	87.99	58	57.99
1	36	35.99	18	17.99	34	33.99
2	10	9.93	11	9.14	14	16.17
3	2	2.36	4	4.79	8	7.42
4	1	0.73	1	2.09	6	3.41
5					2	3.02
Total & χ^2	122	122.00, 0.15	122	122, 1.07	122	122, 2.71

Concluding remark: Thus this paper defined and studied a class of hermite type Lagrangian probability distribution, by well known Lagrange's expansion, with application to various fields of experiments. The fitting of LHPD-II will be investigated latter on.

References

1. P.C. Consul and L.R. Shenton, Use of Lagrange expression for generating generalized probability distribution. SIAM Journal of Applied Mathematics. **23** 239-248 (1972).
2. P.C. Consul and L.R. Shenton, Some interesting properties of Lagrangian distributions. Communications in Statistics. **2** 263-272 (1973).
3. P.C. Consul and L.R. Shenton, On the probabilistic structure and properties of discrete Lagrangian distributions. Statistical Distribution in Scientific Work I: Models and Structure, (Edited by G.P.Patil, S.Kotz and J.K.Kotz), 41-48, Dordrecht Reidel, (1975)
4. S.G. Mohanty, On generalized two-coin tossing problem, Biometrische Zeitschrift, **8** 266-272 (1966).
5. P.C. Consul and G.C. Jain, A generalization of Poisson distribution, Technometrics, **15** 791-799 (1973).
6. P.C. Consul, Lagrange and related probability distribution, Encyclopedia of Statistical Sciences, (Edited by S. Kotz, N.L. Johnson and C.B. Read), **4** 448-454, New York: Wiley, (1983)
7. K.G. Janardan and B.R. Rao, Lagrange distribution of second kind and weighted distributions, SIAM Journal of Applied Mathematics, **43** 302-313, (1983).
8. C.D. Kemp and A.W. Kemp, Some properties of Hermite distribution, Biometrika, **52** 381-394, (1966)
9. M.G. Kendall, Natural law in the social sciences, Journal of Royal Statistical Society, Series **A 124** 1-9, (1961).
10. C.B. Williams, The number of publications written by biologists, Annals of Eugenics, **12**, 143-146, (1944).
11. G.C. Jain, An power series distribution associated with Lagrange expression, Biometrische Zeitschrift, **17**, 85-97, (1975).
12. A.M. Adelstein, Accidents proness a critism of the concept based upon analysis of shunter's accidents, Journal of Royal Statistical Society, Series **A**, **115**, 345-410. (1952)

Jour. Ind. Soc. Ag. Statistics
54(3), 2001 : 317-323

A Study on the Inflated Poisson Lindley Distribution

M. Borah and A. Deka Nath[1]
Tezpur University, Tezpur
(Received : June, 2000)

SUMMARY

The Poisson Lindley distribution has been further studied with some inflation of probability at zero. Some properties of this Inflated Poisson Lindley (IPL) distribution are discussed. The recurrence relations are obtained without derivatives, so that they will be easy to handle on computer for computation of higher order probabilities, moments, etc. The parameters of this distribution have been estimated by three methods. Examples are given for fitting of this distribution to real data, and the fit is compared with that obtained by using other distributions.

Key words : Poisson-Lindley distribution, Inflated distribution, Recurrence relation, Raw moments, Skewness, Kurtosis, Parameter estimation.

1. Introduction

Poisson Lindley distribution is a generalized poisson distribution (see Consul [5]) originally due to Lindley [10] with probability mass function

$$P_x(\phi) = \frac{\phi^2 (\phi + 2 + x)}{(\phi + 1)^{x+3}} \qquad x = 0, 1, 2, \ldots \qquad (1.1)$$

Sankaran [12] further investigated this distribution with application to errors and accidents. In both the examples, single parameter Poisson Lindley distribution gives a better fit than Poisson distribution. It is a special case of Bhattacharya's [2] more complicated mixed poisson distribution. Some mixture of Poisson Lindley distributions derived by using Gurland's generalization [7] were studied by Borah and Deka Nath [4], where certain properties of Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distributions were investigated.

A random variable X is said to have the discrete inflated distribution if its probability function is given by

[1] Department of Statistics, Darrang College, Tezpur, Assam

$$P(X = x) = \begin{cases} \omega + (1 - \omega) p_0 & x = 0 \\ (1 - \omega) p_x. & x = 1, 2, 3, \ldots \end{cases} \tag{1.2}$$

where ω is a parameter assuming arbitrary values in the interval (0, 1). It is also possible to take $\omega < 0$, provided $\omega + (1 - \omega) p_0 \geq 0$ (See Johnson *et al.* [9]).

The discrete inflated distribution was first investigated by Singh [15]. He studied inflated poisson distribution to serve the probabilistic description of an experiment with a slight inflation at a point, say zero. Later Singh ([13], 14]) pointed out that there exists analogous situations in binomial distribution, i.e. distinct increase of frequency of observed event at point zero as well as respective decrease of its value at the remaining points. Pandey [11] studied the generalized inflated poisson distribution. Gerstenkorn [6] established the recurrence relation for the moments for the inflated negative binomial, poisson and geometric distribution.

In this paper, an Inflated Poisson-Lindley (IPL) distribution is discussed to serve the probabilistic description of an experiment with a slight inflation of probability at zero. The recurrence relations for moments and probabilities for IPL distribution are obtained. For fitting of the IPL distribution, three well-known data sets are considered for an empirical comparison and it is observed that this distribution gives better fit in all the cases.

2. *Recurrence Relation for Probabilities*

The probability generating function (p.g.f.), G(t) of IPL distribution may be written as

$$G(t) = \omega + (1 - \omega) g(t) \tag{2.1}$$

where $g(t) = \dfrac{\phi^2 (\phi + 2 - t)}{\{(\phi + 1)(\phi + 1 - t)^2\}}$ is the p.g.f. of Poisson Lindley (PL) distribution $0 < \omega < 1, \phi > 0$ (see Sankaran [12]). Differentiating (2.1) w.r.t. 't' and equating the coefficients of t^r from both sides, we have

$$P_r = \frac{(\phi + 2 + r)}{(\phi + 1)(\phi + 1 + r)} P_{r-1} \qquad r > 1 \tag{2.2}$$

where $P_0 = \omega + (1 - \omega)\phi^2 (\phi + 2)/(\phi + 1)^3$ and $P_1 = \dfrac{(1 - \omega)\phi^2 (\phi + 3)}{(\phi + 1)^4}$

3. Recurrence Relation for Moments

The raw moments recurrence relation for IPL distribution may similarly be written as

$$\mu_r' = \frac{(1-\omega)\{(\phi+3)-2^r\}}{\phi(\phi+1)} + \sum_{j=0}^{r-1} \frac{\left(3\alpha - 3\times 2^{j+1}\alpha^2 + 2^{j+1}\alpha^3\right)}{(1-\alpha)^3}\binom{r}{j+1}\mu_{r-j}', \; r>1 \quad (3.1)$$

where $\alpha = \dfrac{1}{(\phi+1)}$

$$\mu_1' = \frac{(1-\omega)(\phi+2)}{\{\phi(\phi+1)\}} \text{ and } \mu_2' = \frac{(1-\omega)(\phi^2+4\phi+6)}{\{\phi^2(\phi+1)\}}$$

Thus the variance may be obtained as

$$\mu_2 = \frac{(1-\omega)\left\{\phi^3+4\phi^2+6\phi+2+\omega(\phi+2)^2\right\}}{\left\{\phi^2(\phi+1)^2\right\}} \quad (3.2)$$

Putting $\omega = 0$ in (3.2) the variance of PL distribution may be obtained (see Borah *et al.* [4]). The expression for the coefficient of skewness and kurtosis can be written in terms of ϕ and ω

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{P}{Q} \quad (3.3)$$

where $P = \{\phi^5 + 7\phi^4 + 22\phi^3 + 32\phi^2 + 18\phi + 4 + \omega(3\phi^4 + 17\phi^3 + 36\phi^2 + 30\phi + 4)$

$+\, \omega^2(\phi^3 + 6\phi^2 + 12\phi + 18)\}$

and $Q = \sqrt{(1-\omega)\left\{\phi^3+4\phi^2+6\phi+2+\omega(\phi+2)^2\right\}^3}$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{A + \omega B + 3\omega^2 C + 3\omega^3 D}{(1-\omega)\left\{\phi^3+4\phi^2+6\phi+2+\omega(\phi+2)^2\right\}^2} \quad (3.4)$$

where $A = \phi^7 + 2\phi^6 + 73\phi^5 + 174\phi^4 + 256\phi^3 + 152\phi^2 - 24\phi + 12$

$B = 7\phi^6 + 54\phi^5 + 181\phi^4 + 312\phi^3 + 34\phi^2 + 264\phi + 12$

$C = 4\phi^5 + 30\phi^4 + 62\phi^3 + 112\phi^2 + 32\phi$ and

$D = 2\phi^4 + 16\phi^3 + 52\phi^2 + 64\phi + 32$

It is clear from the above expression of γ_1 that for any given value of $\phi > 0$ and ω closes to unity, the skewness is infinitely large and it becomes smaller and smaller as the value of ω decreases. The IPL distribution is easily seen to be leptokurtic as the value of γ_2 is positive for all values of $\phi > 0$ and $0 < \omega < 1$ though there is a factor '-24ϕ' in the numerator of (3.4).

4. *Estimation of Parameters*

The estimation of parameters of inflated distributions other than ω can be carried out by ignoring the observed frequency in the zero class, and then using a technique appropriate to the original distribution truncated by omission of zero class. After the other parameters have been estimated, parameter ω can then be estimated by equating the observed and expected frequencies in the zero class (See Johnson *et al.* [9]). Three methods for estimating parameters of IPL distribution, i.e. method of maximum likelihood, method of moments and ratio of first two frequencies with mean are discussed in this section.

(a) Method of Maximum Likelihood (ML): Since IPL distribution is a zero modified distribution, one of the ML equations is (see Johnson *et al.* [9])

$$\hat{\omega} + \frac{(1-\hat{\omega})\hat{\phi}^2(\hat{\phi}+2)}{(\hat{\phi}+1)^3} = \frac{n_0}{N} \tag{4.1}$$

where $\frac{n_0}{N}$ is the observed proportion of zeros. It is also a power series distribution so the other ML equation will be

$$\bar{x} = \frac{(1-\hat{\omega})(\hat{\phi}+2)}{\{\hat{\phi}(\hat{\phi}+1)\}} \tag{4.2}$$

Eliminating $\hat{\omega}$ from equation (4.1) and (4.2), we have

$$\frac{\hat{\phi}(\hat{\phi}+1)}{(\hat{\phi}+2)}\bar{x} - \frac{\hat{\phi}^2}{(\hat{\phi}+1)^2} = 1 - \frac{n_0}{N} \tag{4.3}$$

$\hat{\phi}$ can be estimated from equation (4.3) by using Newton Raphson method and then $\hat{\omega}$ may be estimated from equation (4.1).

(b) Methods of Moments : The parameters may be obtained from the moments as

$$\hat{\phi} = \frac{\left\{(2\mu_1' - \mu_2') + \sqrt{(\mu_2' - 2\mu_1' + 2\mu_1'\mu_2')}\right\}}{(\mu_2' - \mu_1')} \tag{4.4}$$

$$\hat{\omega} = 1 - \frac{\hat{\phi}(\hat{\phi}+1)\mu_1'}{(\hat{\phi}+2)} \tag{4.5}$$

where μ_1' and μ_2' denote mean and second order raw moments respectively.

(c) Ratio of First Two Frequencies and Mean: Eliminating ω between first two frequencies, we get

$$\hat{\phi} = \left(\frac{n_1}{2n_2} - 2\right) + \sqrt{\left\{\left(2 - \frac{n_1}{2n_2}\right)^2 - \left(3 - \frac{4n_1}{n_2}\right)\right\}} \quad (4.6)$$

where $\frac{n_1}{N} = \frac{(1-\omega)(\hat{\phi}+3)\hat{\phi}^2}{(\hat{\phi}+1)^4}$ and $\frac{n_2}{N} = \frac{(1-\omega)(\hat{\phi}+4)\hat{\phi}^2}{(\hat{\phi}+1)^5}$, are the first two relative frequencies and $\hat{\omega}$ may be estimated from equation (4.5).

5. *Fitting of IPL Distribution to Data*

For the fitting of IPL distribution, we consider two data sets of Beall [1] in Tables 1 and 2, for which generalized Poisson distribution (GPD) was fitted by Jain [8] (using MLE). In Table 3, we consider Student's historic data on Haemocytometer of yeast cells, for which Gegenbauer distribution was fitted by Borah [3], using method of moments. It is observed from Table 1, 2 and 3 that ML gives better result in all the cases. In case of Table 2 the method ratio of first two frequency with mean does not give better fit, as the computed χ^2 value is quite large, hence the result is not reported in this case. It is also clear from the values of the expected IPL frequencies that there is some improvement, however small it may be, in fitting of IPL distribution over the other distributions considered earlier.

Table 1. Fit of distribution on Pyrausta nublilalis in 1937 (data of Beall [1])

No. of Insects	Observed Frequency	IPL (Maximum Likelihood)	IPL (Method of Moments)	IPL (Ratio of Two Freq.)	GPD (Jain [8])
0	33	33.00	32.07	34.08	32.46
1	12	12.41	13.47	11.23	13.47
2	6	5.84	6.00	5.61	5.60
3	3	2.66	2.59	2.71	2.42
4	1	1.18	1.096	1.28	1.08
5	1	0.91	0.774	1.09	0.97
Total	56	56.00	56.00	56.00	56.00
Parameter estimates		$\hat{\phi} = 1.588$ $\hat{\omega} = 0.1406$ $\chi^2 = 0.029$	$\hat{\phi} = 1.719$ $\hat{\omega} = 0.0573$ $\chi^2 = 0.215$	$\hat{\phi} = 1.449$ $\hat{\omega} = 0.228$ $\chi^2 = 0.096$	$\chi^2 = 0.25$

Table 2. Fit of distribution of Corn Borer (data of Beall [1])

Corn Borer per Hill	Observed Frequency	IPL (Maximum Likelihood)	IPL (Method of Moments)	GPD (Jain [8])
0	43	42.99	44.99	43.91
1	35	32.12	30.39	32.00
2	17	19.45	18.81	19.11
3	11	11.31	11.19	10.88
4	5	6.40	6.47	6.12
5	4	3.55	3.66	3.44
6	1	1.94	2.04	1.94
7	2	1.05	1.12	1.10
8	2	1.19	1.30	1.50
Total	120	120	120	120
Parameter estimates		$\hat{\phi} = 1.0587$ $\hat{\omega} = -0.5696$ $\chi^2 = 0.577$	$\hat{\phi} = 1.0715$ $\hat{\omega} = -0.0087$ $\chi^2 = 0.995$	$\chi^2 = 0.87$

Table 3. Haemocytometer Counts of Yeast Cells

No. of Yeast cells per sq.	Observed Frequency	IPL (Maximum Likelihood)	IPL (Method of Moments)	IPL (Ratio of first Two Freq)	Gegenbauer (Borah [3])
0	213	213.00	210.46	204.00	214.15
1	128	127.00	131.14	139.18	123.00
2	37	40.91	40.76	40.23	44.88
3	18	12.82	12.39	11.39	13.36
4	3	3.95	3.71	3.18	3.55
5	1	1.20	1.09	0.88	0.86
6	0	0.53	0.45	0.34	0.20
Total	400	400.00	400.00	400.00	400.00
Parameter estimates		$\hat{\phi} = 2.669$ $\hat{\omega} = -0.431$ $\chi^2 = 1.037$	$\hat{\phi} = 2.774$ $\hat{\omega} = -0.497$ $\chi^2 = 1.53$	$\hat{\phi} = 3.0328$ $\hat{\omega} = -0.6586$ $\chi^2 = 3.93$	$\chi^2 = 2.8342$

ACKNOWLEDGEMENT

The authors are grateful to the referee for his valuable comments which improved the paper considerably.

REFERENCES

[1] Beall, G. (1940). The fit and significance of contagious distributions when applied to observations on larval insects. *Ecology*, **21**, 460-474.

[2] Bhattacharya, S.K. (1966). Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness. *Bulletin of the Cal. Stat. Assoc.*, **15**, 20-31.

[3] Borah, M. (1984). The Gegenbauer distribution revisited: Some recurrence relations for moments, cumulants, etc., estimation of parameters and its goodness of fit. *Jour. Ind. Soc. Agril. Stat.*, **36**, 72-78.

[4] Borah, M. and Deka Nath, A. (2001). Poisson Lindley and some of its mixture distributions. *Pure and Applied Mathematika Sciences*, **53**, 1–2, 1–8.

[5] Consul, P.C. (1989). *Generalized Poisson Distribution Properties and Applications*. Marcel Dekker, Inc., New York and Basel.

[6] Gerstenkorn, T. (1979). Moment recurrence relation for the inflated negative binomial, poisson and geometric distribution. *Demonstratio Mathematica*, **11(2)**, 389-411.

[7] Gurland, J. (1957). Some interrelations among compound and generalized distributions. *Biometrika*, **44**, 265-268.

[8] Jain, G.C. (1974). On power series distribution association with lagrange expansion. *Biometrische Zeitsehrift*, **17**, 85-97.

[9] Johnson, N.L, Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions*. 2nd Edn, John Wiley & Sons Inc.

[10] Lindley, D.V. (1958). Fiducial distributions and Baye's theorem. *J. Roy. Statist. Soc.*, **B 20**, 102-107.

[11] Pandey, K.N. (1965). On generalized inflated poisson distribution. *Banaras Hindu Univ. J. Sci. Res.*, **15**, 157-162.

[12] Sankaran, M. (1970). The discrete poisson lindley distribution. *Biometrics*, **26**, 145-146.

[13] Singh, M.P. (1965). Inflated binomial distribution. *Banaras Hindu Univ. J. Sci. Res.*, **16(1)**, 87-90.

[14] Singh, M.P. (1966). A note on generalized inflated binomial distribution. *Sankhya*, **A28 (1)**, 99.

[15] Singh, S.N. (1963). A note on inflated poisson distribution. *J. Ind. Statist. Assoc.*, **1**, 140-144.

Pure And Applied Mathematika Sciences
Vol. LIII, No. 1-2, March 2001

POISSON-LINDLEY AND SOME OF ITS MIXTURE DISTRIBUTIONS

M. Borah and A. Deka Nath

Abstract

The discrete Poisson-Lindley distribution is a one-parameter mixture distribution obtained from Poisson distribution by mixing with one due to Lindley. In this paper an attempt has been made to review some of the properties like recurrence relations for probabilities, moments etc and to study the problem of estimation for the fitting of the Poisson-Lindley distribution to some well known data. Two generalized distributions namely Poisson-Poisson-Lindley and Poisson Lindley-Poisson are also investigated. The recurrence relations with out any derivatives have been obtained for the computation of higher order probabilities and factorial moments of the above newly derived distributions. The parameters of the distributions have been estimated in terms of first two moments, and also in terms of mean and ratio of first two frequencies. A few sets of reported data, to which different types of 'derived' distributions are fitted with varied amount of success, have been considered for the fitting of the above distributions

1. Introduction

Poisson-Lindley distribution is a mixture distribution obtained by mixing the Poisson distribution with one due to Lindley (1958). Sankaran (1970) further studies this distribution with applications to errors and accidents. Some of the difficul ies in obtaining the MLE of the parameter θ of this single parameter distribution is pointed out by him, and two

Key Words : Poisson-Lindley, Mixture of Distributions, Recurrence Relation, Factorial Moments, and Parameter Estimation

applications to the data suggested that the present distribution can be used as an approximation to the negative binomial (1920) and the hermite distribution (1965).

In this paper Poisson-Lindley distribution is further investigated. Recursive relationship of the probabilities and factorial moments are studied. Two mixture distribution of Poisson-Lindley distribution obtained by mixing Poisson distribution with Poisson-Lindley distribution and Poisson-Lindley mixture of Poisson distribution are also investigated. Recurrence relations for factorial moments and probabilities are also discussed. The aim of this paper is to derive some basic properties of both of these three distributions and to compare it with other distributions on the basic of their fits to empirical data.

2. Poisson-Lindley distribution

(a) Expression for probabilities :

The probability generating function (pgf) of Poisson-Lindley distribution is

$$G(t)=\theta^2\,(\theta+2-t)/(\theta+1)\,(\theta+1-t)^2. \qquad \ldots(2.1)$$

Differentiating the pgf with respect to t the following recurrence relation for probabilities may be written as

$$P_r=[(\theta+2+r)/(\theta+1)\,(\theta+1+r)]\,P_{r-1} \qquad \ldots(2.2)$$

$$P_0=\theta^2\,(\theta+2-t)/(\theta+1)^3 \qquad \text{(see Sankaran (1970))}.$$

Putting $r=1, 2, 3\ldots$ in equation (2.2), the higher order probabilities may be computed easily.

(b) Factorial Moments :

The moment generating function of Poisson-Lindley distribution is given as

$$G(t+1)=\{1-t/(\theta+1)\}/(1-t/\theta)^2. \qquad \ldots(2.3)$$

On differentiating (2.3) w.r.t. and computing coefficient of t^r, on both sides of the equation we obtaino

$$\mu'_{(r+1)}=(r+1)\left[2\mu'_r-r\mu'_{(r-1)}/\theta\right], \qquad \text{for } r=2, 3, 4, \ldots \qquad \ldots(2.4)$$

where

$$\mu'_{(1)}=(\theta+2)/\{\theta(\theta+1)\},$$

and

$$\mu'_{(2)}=2!\,(\theta+3)/\{\theta^2(\theta+1)\}$$

$\mu'_{(r)}$ stands for the rth factorial moments.

Explicit expression for the 3rd and 4th order factorial moments may be written as

$$\mu'_{(3)} = 3!\,(\theta+4)/\{\theta^3(\theta+1)\}$$

$$\mu'_{(4)} = 4!\,(\theta+5)/\{\theta^4(\theta+1)\}.$$

Variance $= \mu = (\theta^3+4\theta^2+6\theta+2)/\{\theta^2(\theta+1)^2\}$.

(c) **Estimation :**

The single parameter 'θ' of the Poisson-Lindley distribution can be estimated in the following methods :

(1) **Method of moments :**

The parameter θ of Poisson-Lindley distribution is estimated by Sankaran (1970) as

$$\theta^* = \left[\left(\mu'_1 - 1\right) + \sqrt{\left\{\left(\mu'_1 - 1\right)^2 + 8\mu'_1\right\}}\right] \Big/ 2\mu'_1 .$$

Where μ'_1 denotes the mean of the distribution.

(2) **Ratio of first two frequencies and the mean :**

For Poisson-Lindley distributions, θ may be estimated by taking ratio of first two frequencies

$$\theta = [-(3f_1 - f_0) + \sqrt{\{(3f_1 - f_0)^2 - 4f_1(2f_1 - 3f_0)\}}]/2f_1,$$

where

$$P_0 = f_0/N = \{\theta^2(\theta+2)\}/(\theta+1)^3,$$

and

$$P_1 = f_1/N = \{\theta^2(\theta+3)\}/(\theta+1)^4.$$

3. **Poisson-Poisson-Lindley distribution**

Poisson-Poisson-Lindley distribution may be derived by generalizing Poisson distribution [see Gurland (1957)], using Poisson-Lindley as generalizing distribution.

(a) **Expression for probability :**

The pgf of Poisson mixture of Poisson-Lindley distribution may be written as

$$G(t) = \mathrm{EXP}\,[\lambda\{\theta^2(\theta+2-t)/(\theta+1)(\theta+1-t)^2 - 1\}]. \qquad \ldots(3.1)$$

The probability recurrence relation may be written as

$$P_{r+1} = [\{3\alpha r + \lambda\theta^2\{2\alpha(2+\theta) - 1\}/(\theta+1)^3\}\,P_r - \{3\alpha^2(r-1) + \lambda\theta^2\alpha/(\theta+1)^3\}\,P_{r-1} + \alpha^3(r-2)\,P_{r-2}]/(r+1) \qquad \ldots(3.2)$$

$$\lambda > 0 \text{ and } \theta > 0.$$

Where

$$P_0 = \mathrm{Exp}\,[\lambda\{\theta^2(\theta+2)/(\theta+1)^3 - 1\}],$$

$$P_1 = [\lambda\theta^2\{2\alpha(2+\theta) - 1\}/(\theta+1)^3]\,P_0,$$

and

$$P_2=[3\alpha P_1+\lambda\theta^2\{\{2\alpha(2+\theta)-1\}P_1-\alpha P_0\}/(\theta+1)^3]/2.$$

(b) Factorial Moments :

The factorial moment generating function for Poisson-Poisson-Lindley distribution may be written as

$$G(1+t)=\text{EXP}\,[\lambda\{\{1-t/(\theta+1)\}/(1-t/\theta)^2-1\}]. \qquad \ldots(3\ 3)$$

Hence the factorial moments recurrence relation

$$\mu'_{(r+1)}=\{3r/\theta+\lambda(\theta+2)/\theta(\theta+1)\}\,\mu'_r-\{3r(r-1)/\theta^2+\lambda r/\theta(\theta+1)\}\,\mu'_{(r-1)}$$

$$+r(r-1)(r-2)\,\mu'_{(r-2)}/\theta^3. \qquad \ldots(3.4)$$

$$\text{Mean}=\mu'_{(1)}=\lambda(\theta+2)/\theta(\theta+1),$$

and

$$\mu'_{(2)}=\lambda^2(\theta+2)^2/\theta^2(\theta+1)^2+2\lambda(\theta+3)/\theta^2(\theta+1),$$

$$\mu'_{(3)}=\lambda^3(\theta+2)^3/\theta^3(\theta+1)^3+2\lambda^2(\theta+2)(5\theta+9)/\theta^3(\theta+1)^2$$
$$+6\lambda(2\theta+8)/\theta^3(\theta+1),$$

$$\mu'_{(4)}=\lambda^4(\theta+2)^4/\theta^4(\theta+1)^4+\lambda^3(\theta+2)^2(7\theta+27)/\theta^4(\theta+1)^3+12\lambda^2(7\theta^2$$
$$+28\theta+29)/\theta^4(\theta+1)+12\lambda(11\theta+28)/\theta^4(\theta+1).$$

$$\text{Variance}=\mu_2=\lambda(\theta^2+4\theta+6)/\theta^2(\theta+1)^2.$$

(c) Estimation :

The two parameter 'λ, θ' of Poisson-Poisson-Lindley distribution can be estimated in the following methods.

(1) Methods of moments :

The two parameters λ, θ of Poisson Poisson-Lindley distribution may be estimated by using sample mean and variance

$$\theta^*=[-(2\bar{x}-S^2)+\sqrt{\{(2\bar{x}-S^2)-6\bar{x}(\bar{x}-S^2)\}}]/(\bar{x}-S^2),$$

and

$$\lambda^*=\bar{x}\{\theta(\theta+1)\}/(\theta+2).$$

Where $\bar{x}$ is the sample mean S^2 is the sample variance.

(2) Ratio of first two frequencies and the mean :

For Poisson-Poisson-Lindley distribution

$$\lambda^*=\{f_1(\theta^*+1)^4\}/f_0\,\theta^{*2}(\theta^*+1)^4,$$

$$\theta^*=[-(2x-S^2)+\sqrt{\{2x-S^2)-6x(x-S^2)\}}]/(x-S^2),$$

where

$$f_0/N=\text{EXP}\,[\lambda\{\theta^2(\theta+2)/(\theta+1)^3-1\}],$$

and

$$f_1/N=[\lambda\theta^2\{2\alpha(2+\theta)-1\}/(\theta+1)^3]\,f_0/N$$

4. Poisson-Lindley-Poisson distribution

Poisson-Lindley-Poisson distribution may be derived by generalising Poisson-Lindley distribution [see Gurland (1957)] using Poisson distribution as generalising distribution.

The pgf of Poisson-Lindley-Poisson can be written as

$$G(t)=[\theta^2 (\theta+2-e^{\lambda(t-1)})]/[(\theta+1)(\theta+1-e^{\lambda(t-1)})^2] \qquad \ldots(4.1)$$
$$=A\,[\theta+2-e^{\lambda(t-1)}]/[1-\alpha\, e^{\lambda(t-1)}]^2,$$

where $A=\theta^2/(\theta+1)^3$ and $\alpha=1/(\theta+1)$.

Hence the probability recurrence relation can be written as

$$P_{r+1}=\Big[A\{(\theta+3)/(\theta+1)\,\lambda e^{-\lambda}-2^r\,\lambda e^{-2\lambda}\}\,\lambda^r/r!+3\alpha e^{-\lambda}\Big[\sum_{j=1}^{r}(1-2^j\,\alpha e^{-\lambda}+3^{j-1}\,\alpha^2\,e^{-3\lambda})\,\lambda^j/j!\,(r-j+1)\,P_{r-j+1}\Big]\Big]\Big/B(r+1), \qquad \ldots(4.2)$$

where

$$B=1/(1-3\alpha e^{-\lambda}+3\alpha^2\,e^{-2\lambda}-\alpha^3\,e^{-3\lambda}),$$
$$P_0=A\,(\theta+2-e^{-\lambda})/(1-\alpha e^{-\lambda})^2,$$
$$P_1=A\lambda\,e^{-\lambda}\,\{(\theta+3)/(\theta+1)-\lambda e^{-\lambda}\}/B.$$

(b) **Factorial moments:**

The factorial moments generating function of Poisson-Lindley-Poisson distribution may be written as

$$G(t+1)=A\,(\theta+2-e^{-\lambda t})/(1-\alpha e^{\lambda t})^2. \qquad \ldots(4.3)$$

The factorial moment recurrence relation may be written as

$$\mu'_{(r+1)}=\Big[A\{(\theta+3)/(\theta+1)\,\lambda^{r+1}-2^r\,\lambda^{r+1}\,\alpha\}+\sum_{j=1}^{r}(3\alpha-3\alpha^2\,2^j+\alpha^3\,3^j)\,{}^{r}c_j\,\lambda^j\,\mu'_{(r-j+1)}\Big]\Big/(1-\alpha)^3. \qquad \ldots(4.4)$$

Putting $r=1, 2, 3, \ldots$ in equation (4.4) respectively, the higher order moments may be obtained as

$$\mu'_{(2)}=\lambda^2\,(\theta^2+4\theta+6)/\theta^2\,(\theta+1),$$

$$\mu'_{(3)}=\lambda^3\,(\theta^3+8\theta+12\theta+48)/\theta^3\,(\theta+1),$$

$$\mu'_{(4)}=\lambda^4\,(\theta^4+16\theta^3+78\theta^2+60\theta+336)/\theta^4\,(\theta+1).$$

Hence mean and variance for Poisson-Lindley-Poisson distribution will be

Mean $-\mu_1' = \lambda(\theta+2)/\theta(\theta+1)$,

and

Variance $=\lambda^2(\theta^3-4\theta^2+6\theta+2)/\theta^2(\theta+1)+\lambda(\theta+2)/\theta(\theta+1)$.

(c) **Estimation :**

The two parameters λ and θ of Poisson-Lindley-Poisson distribution may be estimated as follows.

Let

$$F(\theta)=(\theta^3+4\theta^2+6\theta+2)/(\theta+2)^2-\left(\mu_2-\mu_1'\right)\Big/\mu_1' . \qquad \ldots(4.5)$$

The parameter θ may be estimated by Newton Raphson method using equation (4.5) and the other parameter λ may be estimated as

$\lambda=\bar{X}\,\theta(\theta+1)/(\theta+2)$, where $\bar{x}$ denotes the mean of the distribution.

Goodness of fit :

All these three distributions *i.e.* Poisson-Lindley, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson were fitted to distribution of mistakes in groups of random digits, data from Kemp and Kemp (1965) and accidents to 647 women on high explosive shells in 5 week data from Greenwood and Yule (1920) reported by Kendal and Stuart (1963) for which single parameter Poisson, two parameter Hermite and Negative binomial have been fitted. Since obtaining maximum likelihood estimates is very cumbersome, method of moments are used to estimate the parameters of these distributions. The Table 1 and 2 give the comparison of observed and expected frequencies for these distributions, the Poisson-Lindley, the Poisson-Poisson-Lindley distribution and the Poisson-Lindley-Poisson distribution.

Table 1. Observed and fitted Poisson-Lindley, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distributions.

No. of Accident	Observed Frequency	Expected Frequency				
		Poisson	NB	PL	PPL	PLP
0	447	406	441	439.28	442.05	444.58
1	132	189	140	142.83	137.79	134.78
2	42	45	45	45.02	46.57	46.17
3	21	7	14	13.88	14.57	14.85
4	3	1	5	4.20	4.32	4.62
≥ 5	2	1	2	1.79	1.7	2 00
Total	647	649	648	647.00	647.00	647.00
χ^2				3.05	2.17	1.40
Degrees of Freedom				4	3	3

Table 2. Observed and fitted Poisson-Lindley, Poisson-Poisson-Lindley and Poisson-Lindley-Poisson distributions.

No. of errors per group	Observed Frequency	Expected Frequency				
		Poisson	N. Bio.	PL	PPL	PLP
0	35	27.4	34.2	33.05	32.83	35.53
1	11	21.5	11.7	15.27	15.22	15.69
2	8	8.4	9.6	6.74	7.06	7.02
3	4	2.2	2.8	2.89	2.99	2.91
4	2	.4	1.3	1.21	1.19	1.15
Total	60	59.9	59.6	59.17	59.29	59.29
χ^2				2.23	1.99	2.179
Degrees of freedom				3	2	2

Form the a ove table it is clear that there is some improvement, however small it may be, in fitting these mixture distributions PL, PPL and PLP over the other distributions consider earlier. The distributions as indicated here, may be used with case in the other situations also.

REFERENCES

[1] Bhattacharya, S. K., Confluent hypergeometic distributions of discrete and Continuous type with applications to accident proneness, *Bulletin of the Calcutta Statistical Association*, 15, 20–31 (1966).

[2] Bora, M., The Gegenbenbauer Distribution Revisited : Some recurrence Relations for Meoments, Cumulant, etc., Estimation of Parameters and its goodness of fit, *Journal of the Indian Society of Agricultural Statistics*, New Delhi, Vol. 34, No. 1, 72-78 (1984).

[3] Greenwood, M. and Yule, G. U., An inquiry into the nature of frequency distributions representative of multiple happnengs with particular reference to the occurrence of multiple attacks of disease or of repeated accidents; *Journal of the Royal Statistical Society*, series A, 83, 255-279 (1920).

[4] Gurland, J., Some interrelations among compound and generalized distributions, *Biometrika*, 44, 265-268 (1957).

[5] Kemp, C. D. and Kemp, A. W., Some properties of the Hermite Distribution, *Biometrics*, 52, 381-394 (1965).

[6] Lindley, D. V., Fiducial distributions and Baye's theorem, *Journal of the Royal Statistical Society*, Series B, 20, 102-107 (1958).

[7] Medhi, J. and Bora, M., On generalized Four Parameter Charlier Distribution *Journal of Statistical Planning and Inference*, 69-77 (1986).

[8] Sankaran, M., The discrete Poisson-Lindley distribution, *Biometrics*, 26, 145–146 (1970).

Received : September 25, 1999
Department of Mathematical Sciences,
Tezpur University, Tezpur—784001, Assam.

and

Department of Statistics,
Darrang College, Tezpur, Assam.

Journal of Asssam Science Society, Vol.41 No.2, June, 2000, pp. 120-128

The Short Poisson-Poisson-Lindley Distribution

Arundhati Deka Nath
Department of Statistics
Darrang College, Tezpur, Assam, India

and

M. Borah
Department of Mathematical Sciences
Tezpur University, Tezpur, Assam, India

ABSTRACT

The Short Poisson-Poisson-Lindley (SPPL) distribution is an extension of Poisson-Poisson-Lindley (PPL) distribution. It is a convolution of PPL and Poisson distribution. This convolution has been made by assuming that the number of spells in a given time period is assumed to be Poisson variable and the probability of accidents within a spell have a Poisson-Lindley distribution, which is a more generalized Poisson distribution with a constant parameter and accidents occurring outside the spell are independently distributed as Poisson distribution. The recurrence relation for probabilities and factorial moments for this SPPL distribution are discussed. A few sets of accident data, to which the different types of distributions are fitted with varied amount of success, have been considered for fitting of the SPPL distribution.

Key Words : *Short distribution, Poisson-Lindley distribution, Poisson-Poisson-Lindley distribution, recurrence relation, factorial moments, parameter estimation.*

INTRODUCTION

Cresswell and Froggatt (1963) derived a model which was a convolution of Poisson distribution and a Neyman Type A distribution. They called it 'Short', as opposed to the two-parameter

'Long' Neyman Type A distribution. The name 'Short' appears to relate the tails of the distribution. Kemp (1967) considered its properties, recurrence relations for probabilities and fitted this distribution to accident data. This SPPL distribution is also a convolution of Poisson distribution and Poisson-Poisson-Lindley (unpublished work of the authors) distribution. Here we have considered Poisson-Lindley distribution since it is a more generalized Poisson distribution. Sankaran (1970) studied the Poisson-Lindley distribution with applications to errors and accidents. Later two mixture distributions of this distribution namely Poisson-Poisson-Lindley and Poisson-Lindley-Poisson was investigated by Borah and Deka Nath (unpublished work) with application to accident data.

Model Derivation from Accident Data

While deriving the 'Short' distribution from accident data, four assumptions were made by Cresswell and Froggatt (1963). In the same manner, SPPL distribution has been derived by considering the following assumptions :

(i) Every driver is liable to a spell during which he is liable to incur accidents. The number of spells in a given time period is assumed to be Poisson variable with parameter λ_1.

(ii) All drivers are equally liable to the occurrence of a spell.

(iii) The probability of an accident occurring within a spell is constant and it is assumed to have a Poisson-Lindley distribution with constat parameter θ.

(iv) Lastly, accidents can occur outside a spell and such accidents are independently distributed as Poisson distribution with parameter λ_2.

It is generally observed that the derivation of probability mass function (p.m.f.) for some generalized mixture distributions are seem to be complicated. So, the p.g.f. of SPPL distribution has been obtained by using Levy's theorem [see Feller (1957)], which may be written as

$$G(t) = \exp[\{\lambda_1(g(t) - 1) + \lambda_2(t - 1)\}] \quad (1.1)$$

which converges for $|t| \leq 1$, where t is the generating parameter and $g(t) = \frac{\theta^2(\theta + 2 - t)}{(\theta + 1)(\theta + 1 - t)^2}$ denotes the p.g.f. of Poisson Lindley distribution [see Sankaran (1970)]. Hence, (1.1) may be written as

$$G(t) = \exp\left[\lambda_1\left\{\frac{\theta^2(\theta + 2 - t)}{(\theta + 1)(\theta + 1 - t)^2} - 1\right\} + \lambda_2(t - 1)\right] \quad (1.2)$$

be the p.g.f. of SPPL distribution, where $\lambda_1, \lambda_2 > 0$ and $\theta > 0$.

In this paper, we have obtained the recurrence relation for probabilities and factorial moments of the distribution. The parameters are estimated by a composite method i.e. by using the ratio of first two frequencies and first two moments. To illustrate the various applications of this distribution, first we consider the number of accidents sustained by a group of 708 bus drivers over a period of 3 years. Secondly, we condider the number of accidents to 647 women on high explosive shells in 5 week periods. Thirdly, we consider the number of accidents (home injuries) of 122 experienced men during 11 years period and lastly, we considered number of accidents of 122 experienced shunting men over a period of 11 years. In all the cases the SPPL distribution provides a better fit to the observed data.

EXPRESSION FOR PROBABILITIES

Differentiating p.g.f. (1.1) w.r.t. 't', the following recurrence relation for probabilities may be obtained as

$$P_{r+1} = \frac{1}{r+1}\left[\left\{\frac{\lambda_1\theta^2(\theta+3)}{(\theta+1)^4} + \frac{3r}{(\theta+1)} + \lambda_2\right\}P_r - \left\{\frac{\lambda_1\theta^2}{(\theta+1)^4} + \frac{3(r-1)}{(\theta+1)^2} + \frac{3\lambda_2}{\theta+1}\right\}P_{r-1}\right.$$
$$\left. + \left\{\frac{3\lambda_2}{(\theta+1)^2} + \frac{r-2}{(\theta+1)^3}\right\}P_{r-2} - \frac{1}{(\theta+1)^3}P_{r-3}\right], \tag{2.1}$$

where
$$P_0 = exp\left[\lambda_1\left\{\frac{\theta^2(\theta+2)}{(\theta+1)^3} - 1\right\} - \lambda_2\right] \tag{2.2}$$

$$P_1 = \left\{\lambda_1\frac{\theta^2(\theta+3)}{(\theta+1)^4} + \lambda_2\right\}P_0 \tag{2.3}$$

Putting r = 1, 2, 3..... in equation (2.1) the higher order probabilities may be obtained.

EXPRESSION FOR FACTORIAL MOMENTS

The factorial moment generating (f.m.g.) function of the new distribution is

$$G(1+t) = exp\left\{\lambda_1\frac{1-\frac{t}{(\theta+1)}}{\left(1-\frac{t}{\theta}\right)^2} + \lambda_2 t\right\} \tag{3.1}$$

On differentiating (3.1) w.r.t. 't' and comparing the coefficient of t^r, on both sides of the equation, we obtain

$$\mu'_{(r+1)} = \left\{\lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2 + 3\frac{r}{\theta}\right\}\mu'_{(r)} - \left\{\frac{\lambda_1}{\theta(\theta + 1)} + 3\lambda_2 \frac{r}{\theta} + 3\frac{r(r-1)}{\theta^2}\right\}\mu'_{(r-1)}$$

$$+ \left\{3\lambda_2 \frac{r(r-1)}{\theta^2} + \frac{r(r-1)(r-2)}{\theta^3}\right\}\mu'_{(r-2)} - \frac{r(r-1)(r-2)}{\theta^3}\mu'_{(r-3)} \tag{3.2}$$

where $\mu'_{(r)}$ denotes the r^{th} order factorial moments. Explicit expression for the first four factorial moments may be obtained as

$$\mu'_{(1)} = \lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2$$

$$\mu'_{(2)} = \lambda_1^2 \frac{(\theta + 2)^2}{\theta^2(\theta + 1)^2} + 2\lambda_1 \frac{(\theta + 3)}{\theta^2(\theta + 1)} + 2\lambda_1\lambda_2 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2^2$$

$$\mu'_{(3)} = \left\{\lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2 + \frac{6}{\theta}\right\}\mu'_{(2)} - \left\{\lambda_1 \frac{1}{\theta(\theta + 1)} + \frac{6\lambda_2}{\theta} + \frac{3}{\theta^2}\right\}\mu'_{(1)} + 6\frac{\lambda_2}{\theta}$$

$$\mu'_{(4)} = \left\{\lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2 + \frac{9}{\theta}\right\}\mu'_{(3)} - \left\{\lambda_1 \frac{1}{\theta(\theta + 1)} + \frac{9\lambda_2}{\theta} + \frac{18}{\theta^2}\right\}\mu'_{(2)} +$$

$$\left\{18\frac{\lambda_2}{\theta^2} + \frac{6}{\theta}\right\}\mu'_{(1)} + \frac{6}{\theta^3}$$

$$\text{Mean} = \lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2, \qquad \text{Variance} = \lambda_1 \frac{(\theta^2 + 4\theta + 6)}{\theta^2(\theta + 1)} + \lambda_2.$$

If $\lambda_2 \to 0$, then the moments are same as those of Poisson-Poisson-Lindley distribution.

ESTIMATION OF PARAMETERS

A composite method has been used to estimate the parameters of SPPL distribution. The method is based on ratio of first two frequencies, sample mean and sample variance. We have,

$$\bar{x} = \lambda_1 \frac{(\theta + 2)}{\theta(\theta + 1)} + \lambda_2 \tag{4.1}$$

$$S^2 = \lambda_1 \frac{(\theta^2 + 4\theta + 6)}{\theta^2(\theta + 1)} + \lambda_2 \tag{4.2}$$

By equating the first two probabilities of SPPL distribution with $\frac{n_0}{N}$ and $\frac{n_1}{N}$ we obtain from (2.3)

$$\frac{n_1}{n_0} = \lambda_1 \frac{\theta^2(\theta + 3)}{(\theta + 1)^4} + \lambda_2 \qquad (4.3)$$

By eliminating λ_1 and λ_2 between (4.1), (4.2) and (4.3), we obtain

$$\frac{2(\theta + 1)^3(\theta + 3)}{2\theta^4 + 9\theta^3 + 7\theta^2 + 2\theta} = \frac{S^2 - \bar{x}}{\bar{x} - \frac{n_1}{n_0}} \qquad (4.4)$$

which gives an estimate for θ either by graphically or by numerical solution using Newton-Raphson method i.e.,

Let, $$f(\theta) = \frac{2(\theta + 3)(\theta + 1)^3}{2\theta^4 + 9\theta^3 + 7\theta^2 + 2\theta} - K,$$

where $$K = \frac{S^2 - \bar{x}}{\bar{x} - \frac{f_1}{f_0}}$$

Then the iteration formula for Newton-Raphson method is,

$$\theta^* = \theta_0 - \frac{f(\theta)}{f'(\theta)}, \qquad (4.5)$$

where θ_0 is the initial value and θ^* is the estimated value of θ respectively. The initial guess value for starting the Newton-Raphson method have to be selected by trial values, based on our assumptions. When trial value closes to estimated value the method will always convergent.

After getting the estimate of θ i.e. θ^* from (4.5) the estimate of λ_1 and λ_2 may be obtained as

$$\lambda_1^* = \frac{(S^2 - \bar{x})\theta^{*2}(\theta^* + 1)}{2(\theta^* + 3)} \qquad (4.6)$$

$$\lambda_2^* = \bar{x} - \frac{\lambda_1^*(\theta^* + 2)}{\theta^*(\theta^* + 1)} \qquad (4.7)$$

GOODNESS OF FIT

To illustrate the application of this distribution first we consider in Table-1, the data on the number of accidents sustained by a group of 708 bus drivers over a period of three years for which Neyman Type A and Short distribution were fitted by Kemp (1967). When SPPL distribution is applied to this data it provides a surprisingly good fit with χ^2 value of 2.445. Using equations (4.5), (4.6) and (4.7) we get $\theta^* = 8.0110$, $\lambda_1^* = 30.171$ and $\lambda_2^* = -1.892$.

Table-1: Numbers of drivers sustaining accidents over three year period (Cresswell and Forggatt's Table-5.4, 1963) with expected frequencies based on SPPL, Neyman – Type A and short distribution.

No. of accidents	Observed Frequency	SPPL distribution	Neyman-Type A distribution (Kemp, 1967)	Short distribution (Kemp, 1967)
0	117	118.588	116.69	110.38
1	157	159.130	162.04	169.70
2	158	153.191	153.12	156.02
3	115	115.631	115.26	113.90
4	78	74.940	74.58	72.54
5	44	43.182	43.13	41.90
6	21	22.687	22.83	22.45
7	7	11.045	11.25	11.31
8	6	5.041	5.21	5.42
9	1	2.177	2.29	2.49
10	3	0.895	0.96	1.10
11	1	1.493	0.39	0.47
Total	708	708.000	707.75	707.68
χ^2		2.445	2.64	3.78

Secondly, we consider the data on number of accidents to 647 women on high explosive shells during 5 weeks period (data from Greenwood and Yule, 1920). In Table-2 we considered the original data together with the expected frequencies of SPPL, PPL (Borah and Deka Nath, unpublished work) and Negative Bionomial (Plunket and Jain, 1975) distribution. Using equations (4.5), (4.6) and (4.7) we get $\theta^* = 5.3066$, $\lambda_1^* = 2.4117$ and $\lambda_2^* = -0.061$. From Table- 2 we have seen that SPPL distribution provides a good fit to this data.

In Table-3 we consider Adelstein (1952), data on number of accidents (home injuries) of 122 experienced men in 11 years period where Adelstein had concluded that the Poisson distribution fits the first sets of data but did not fit the second and third sets. When the SPPL distribution is applied to those sets it provides a surprisingly good fit in all the cases. In Table-2 we get $\theta^* = 4.2305$, $\lambda_1^* = 0.4401$ and $\lambda_2^* = 0.4171$ for the first set and $\theta^* = 4.0447$, $\lambda_1^* = 3.359$ and $\lambda_2^* = -0.0115$ for the second set. When the estimated value of θ^* are divided by the respective number of years (6 and 11), the average value for θ^* for these sets become 0.705 and 0.3677 respectively. These values do indicate that the average natural rate for home injuries does decrease with experience.

Table-2: Comparison of observed frequencies for accidents to 647 women on high explosive shells during 5 weeks with expected frequencies of SPPL, PPL (Borah and Deka Nath, unpublished work) and Negative Bionomial distribution (Plunket and Jain, 1975). (Data from Greenwood and Yule, 1920).

No. of accident	Observed frequency	SPPL frequency	N. B. frequency	PPL frequency
0	447	445.959	445.89	442.52
1	132	131.692	134.90	137.79
2	43	47.698	44.00	46.57
3	21	15.218	14.69	14.57
4	3	4.124	4.96	4.32
≥ 5	2	2.309	2.56	1.70
Total	647	647.000	647.00	647.00
χ^2		3.199	3.7109	4.041

Table-3: Comparison of observed frequencies for home injuries of 122 experienced men during 11 years (1937 - 1947) with expected SPPL distribution frequencies.

Number of injuries	1937 - 1942		1937 - 1947	
	Observed	Expected	Observed	Expected
0	73	72.952	58	57.045
1	36	35.977	34	33.441
2	10	10.079	14	17.521
3	2	2.313	8	8.161
4	1	0.679	6	3.404
5	-	-	2	2.428
Total	122	122.000	122	122.000
χ^2		0.00068		1.542

In Table-4 we consider accidents data for experienced shunting men over 11 years. For which Adelstein (1952) had used the Poisson distribution and negative binomial distribution. When SPPL distribution is fitted to this data the calculated value of χ^2 for the SPPL distribution is much less than the significant values. In Table-4 we get $\theta^* = 6.7819$, $\lambda_1^* = 6.759$ and $\lambda_2^* = 0.1457$ for the first set and $\theta^* = 1.7748$, $\lambda_1^* = 0.2090$ and $\lambda_2^* = 0.8069$ for the second set. The values of θ^* for the second group indicates that due to experience this group has a less natural chance of making accidents.

Table-4: Comparison of observed frequencies of accidents of 122 experienced shunting men over 11 years (1937 - 1947) with expected SPPL distribution frequencies.

Number of injuries	1937 - 1942		1937 - 1947	
	Observed	Expected	Observed	Expected
0	40	40.141	50	49.642
1	39	39.138	43	49.671
2	26	23.794	17	19.495
3	8	11.459	9	6.731
4	6	4.749	2	2.074
5	2	1.765	0	0.5790
6	1	0.954	1	0.808
Total	122	122.000	122	122.000
χ^2		1.648		1.150

CONCLUSION

It is apparent from the results of the following tables that the SPPL distribution can be applied very successfully in case of accident data. In all the cases the SPPL distribution provides a much better fit than the other distributions.

ACKNOWLEDGEMENT

The authors wish to thank the referees for their thoughtful remarks, which have improved the paper.

REREFENCES

1. Adelstein, A. M. (1952) : Addident proneness : a criticism of the concept based upo analysis of shunter's accidents, ***J. Roy. Stat. Soc. Ser.*** A, 115, 354-410.

2. Cresswell, W. L. and Froggatt, P. (1963) : The causation of bus drivers accidents A epidemivtogical study. ***Oxford University Press,*** London.

3. Greenwood, M. and Yule, G. U. (1920) : An inquiry into the nature of frequenc distributions representative of multiple happenings with particular reference to th occurrence of multiple attacks of dease or of repeated accidents. ***J. of the Roy. Statis Soc.*** Series A, 83, 255-279.

4. Feller, W. (1957) : An introduction to probability theory and its application, ***John Wile & Sons,*** New York.

5. Kemp, C. D. (1967) : On a contagious distribution suggested for accident dat ***Biometrics***, 241-255.

6. Plunkett, I. G. and Jain, G. C. (1975) : Three generalized negative binomial distribution ***Biometrische Zeitschrift***, 17, 276-302.

7. Sankaran, M. (1970) : The Poisson-Lindley distribution. ***Biometrics***, 26, 145-149.

Proceedings of the National Conference on Mathematical and Computational Models,
December 27-28, 2001. PSG College of Technology, Coimbatore, INDIA.

A STUDY ON CHARLIER TYPE LAGRANGIAN PROBABILITY DISTRIBUTIONS

A. Deka Nath[1], M. Borah[2]

[1] Lecturer, Dept. of Statistics
Darrang College

[2] Head of the Dept. Mathematical Sciences
Tezpur University

Tezpur-784001, Assam . Email: dnarundhati@yahoo.com

Abstract: A class of Charlier type Lagrangian probability distributions are defined and studied. The probability mass function (pmf) and cumulants of the basic Lagrangian Charlier distribution is provided. The parameters are estimated by ratio of two moments and first frequency. Fitting of the distribution has been considered for the testing the validity of the estimate of the parameters. Further the general Lagrangian Charlier Poisson distribution of type –I and type –II are also investigated.

Key Words: Lagrange's expansion, Lagrangian probability distribution, Charlier distribution, Probability generating function, Cumulants.

INTRODUCTION

Lagrangian expansions for the derivation of expressions for the probabilities of certain discrete distributions have been used for many years. The potential of this technique for deriving distributions and their properties has been systematically exploited by Consul and Shention and their co-workers. Consul and Shenton [1], [2], [3], Mohanty [4], Consul and Jain [5] have written many key papers on Lagrangian probability distribution. Consul's [6] book on Lagrangian Poisson distribution highlights many properties, and also for various modes of genesis of Lagrangian Poisson distribution.

If g (z) and f (z) are two given probability generating functions (pgf) in 'z' then the transformation z=ug (z) gives Lagrangian probability distribution of type-I, with pmf given by

$$P[X=x]=\frac{1}{x!}\frac{\delta^{x-1}}{\delta z^{x-1}}\left[(g(z))^x\frac{\delta f(z)}{\delta z}\right]_{z=0}, \quad for x=1,2,3,... \qquad (1)$$

and $$P[X=0]=f(0)$$

where $\frac{\delta^{x-1}}{\delta z^{x-1}}$ *denotes* (x-1)th derivative.

Using Lagrangian expansion of 2nd kind Janardan and Rao [7] investigated a new class of discrete distribution call Lagrangian probability distribution of type -II (LPD-II) with pmf

$$P(X=x)=\frac{1-g'(1)}{x!}\left[\frac{\delta^x}{\delta z^x}\{(g(z))^x f(z)\}\right]_{z=0}, \quad for\ x=0,1,2,\dots \tag{2}$$

= 0, otherwise,

where $g'(1)$ denote 1st derivative of g (z) at z=1.

Charlier polynomials defined by the generating function $e^z(1-\beta z)^{-\lambda}$ are associated with the Poisson distribution of rare events. Jain and Gupta [8] defined the generalized Charlier polynomial by the generating function $e^{z\alpha}(1-\beta z^m)^{-\lambda}$.

Medhi and Borah [9] also studied the probability, moments and cumulants properties of four parameter generalized Charlier distribution. The distribution includes, as particular cases, negative binomial, Gegenbauer and generalized Charlier distributions.

Using Lagrange expansion to this Charlier distribution here we have derived the pmf of basic LCD in a simpler form than the earlier one, (See Borah and Begum, [10]) and also the estimation of the parameters. Then considering $f(t)=e^{\theta(z-1)}$ in equation (1), Lagrangian Charlier Poisson distribution of type I and type II (LCPD-I & LCPD-II) are also obtained. The cumulants of the distributions are investigated. For fitting of basic LCD a composite method of estimation of the parameters are suggested. The basic LCD has been fitted to some data for which logarithmic series, geometric and generalized logarithmic series distributions have been fitted. It has been found from Table 1. that this three parameters basic LCD gives a better fit than the other distributions.

1(a). Basic Lagrangian Charlier distribution (LCD)

The probability mass function (pmf) of basic Lagrangian distribution is given as

$$P(X=x)=\frac{1}{x!}\frac{\delta^{x-1}}{\delta z^{x-1}}\{g(z)\}^x\Big|_{z=0}, \text{for } x=1,2,\dots \tag{3}$$

= 0, otherwise.

where g (z) is the probability generating function (pgf) defined on some or all non negative integers, such that $g(0)\neq 0$. Here we consider

$$g(z)=e^{-\alpha}(1-\beta)^{\lambda}e^{\alpha z}(1-\beta z)^{-\lambda} \tag{4}$$

which is the pgf of three-parameter Charlier distribution (for m=1). Thus the pmf of basic LCD may be written as

$$P(X=x)=\frac{e^{-x\alpha}(1-\beta)^{\lambda x}}{x!}\left\{\sum_{j=0}^{k}\binom{k}{j}(\alpha x)^{k-j}\beta^j(\lambda x)_{(j)}\right\}, \text{ for } x=1,2,\dots \tag{5}$$

=0, otherwise.

where k=x-1, $\alpha>0$, $\beta<1$, $\lambda>0$. This pmf may also be written as

$$P(X=x)=\frac{e^{\alpha x}(1-\beta)^{\lambda x}(\alpha x)^{x-1}}{x!}{}_2F_0(1-x,\lambda x,-\beta/\alpha x),\ x\geq 1 \tag{6}$$

(See Borah and Begum [10])

(b) Cumulants of the basic LCD

The cumulants of basic LCD are investigated by using Consul and Shenton's [3] general formula. For simplicity, let G_i be the ith cumulants of the Charlier distribution with pgf g (z) then the first four cumulants k_i, i=1,2,3,4 can be written as

$$k_1=\frac{1}{1-G_1}, \qquad k_2=\frac{G_2}{(1-G_1)^3}, \qquad k_3=\frac{G_3}{(1-G_1)^4}+\frac{3G_2^{\,2}}{(1-G_1)^5},$$

$$k_4=\frac{G_4}{(1-G_1)^5}+\frac{10G_3G_2}{(1-G_1)^6}+\frac{15G_3^{\,2}}{(1-G_1)^7}$$

In case of charlier distribution, we have the first four cumulants as

$$G_1=\alpha+\frac{\lambda\beta}{(1-\beta)}, \qquad G_2=\alpha+\frac{\lambda\beta}{(1-\beta)^2}$$

$$G_3=\alpha+\frac{\lambda\beta(1+\beta)}{(1-\beta)^3} \quad \text{and} \quad G_4=\alpha+\frac{\lambda\beta}{(1-\beta)^4}$$

Thus the first four cumulants of basic LCD may be given as

$$k_1=D_1=\frac{(1-\beta)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}} \tag{7}$$

$$k_2=D_2=\frac{(1-\beta)\{\alpha(1-\beta)^2+\lambda\beta}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3} \tag{8}$$

$$k_3=D_3=\frac{\alpha(1-\beta)^4+\lambda\beta(1-\beta^2)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^4}+\frac{3(1-\beta)\{\alpha(1-\beta)^4+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^5} \tag{9}$$

$$k_4=D_4=\frac{(1-\beta)}{[(1-\alpha)(1-\beta)-\lambda\beta]^5}\left[\{\alpha(1-\beta)^4+\lambda\beta\}+15\frac{\{\alpha(1-\beta)^2+\lambda\beta\}^3}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^2}\right.$$
$$\left.+10\frac{\{\alpha(1-\beta)^3+\lambda\beta(1+\beta)\}\{\alpha(1-\beta)^2+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}}\right] \tag{10}$$

(c) Estimation of Parameters

A composite method has been used to estimate the parameters of basic LCD. By equating the first probability of basic LCD with n_1/N, we obtain

$$\alpha=\lambda\log(1-\beta)-\log(\frac{n_1}{N}) \tag{11}$$

By equating the mean and variance of the basic LCD with $\bar{x}$ and m_2, we get the following equations

$$\bar{x}=k_1=\frac{(1-\beta)}{\{(1-\alpha)(1-\beta)-\lambda\beta\}} \tag{12}$$

$$m_2=k_2=\frac{(1-\beta)\{\alpha(1-\beta)^2+\lambda\beta\}}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3} \tag{13}$$

Eliminating α and λ between Eqs. (11), (12) and (13), we obtain

$$(\frac{1}{\beta}-1)^2\log(1-\beta)+\frac{1}{\beta}=\frac{m_2+\bar{x}^3\log(\frac{n_1}{N})}{m_2+\bar{x}^2-\bar{x}^3} \tag{14}$$

The Eq. (14) may give an estimate for β either by graphically or by using Newton Raphson method. After getting the estimate $\hat{\beta}$ of β from (14) the estimates of λ and α may be obtain as follows,

$$\hat{\lambda} = \{\log(n_1/N) - 1/\bar{x} + 1\} \Big/ \{\log(1-\hat{\beta}) + \hat{\beta}/(1-\hat{\beta})\} \quad (15)$$

and $$\hat{\alpha} = 1 - \frac{1}{\bar{x}} - \frac{\hat{\lambda}\hat{\beta}}{(1-\hat{\beta})} \quad (16)$$

(d) Fitting of basic LCD distribution

For the application of basic LCD, we consider the example of number of paper published per author for which geometric distribution (GD) and logarithmic series distribution (LSD) are fitted by Williams [11] and generalized logarithmic series distribution (GLSD) by Jain [12]. The comparison of observed and expected frequencies among LCD, GLSD, LSD and GD are given in Table 1.

For the data in Table 1, the sample mean $\bar{x}$=1.5508475, and central moments m_2'=1.1405050. Solving Eq. (14) by Newton Raphson method we get $\hat{\beta} = 0.6113$. Substituting the values of $\hat{\beta}$ in Eqs. (15) and (16), we get $\hat{\lambda}$=-0.0199 and $\hat{\alpha}$ =0.3866

Table 1. Fitting of no. papers per author by LCD, GLSD, LSD and GD. Publication in the review of applied entomology, Vol. 24, 1936 (2379 papers by 1534 authors).

No of papers Par author	Observed frequency	L C D frequencies	G L S D Jain [12]	L S D Willams [11]	G D Willams [11]
1	1062	1061.90	1052.72	1046.05	989.10
2	263	275.019	287.52	293.05	351.30
3	120	105.14	107.10	109.46	124.80
4	50	46.083	45.51	45.99	44.33
5	22	22.53	20.83	20.61	15.75
6	7	11.41	10.00	9.62	5.59
7	6	5.76	4.97	4.62	1.99
8	2	3.15	2.53	2.26	0.71
9	0	1.47	1.31	1.12	0.25
10	1	0.43	0.70	0.53	0.09
11	1	0.19	0.81	0.66	0.09
χ^2		4.66	5.14	5.56	46.39

It is clear from table 1 that the expected basic LCD frequencies are much closer to the observed frequencies than obtained by geometric, logarithmic and generalized logarithmic distributions, as the values of χ^2 for the LCD are smaller than for the other distributions. Thus the LCD model better describes the pattern of the frequency distribution of number of paper per author.

2(a). General Lagrangian Charlier type distributions

Considering $g(z)=e^{-\alpha}(1-\beta)^{\lambda}e^{\alpha z}(1-\beta z)^{-\lambda}$ and $f(z)=e^{\theta(z-1)}$ in Eq. (1) the pmf of Lagrangian Charlier Poisson distribution of type I (LCPD-I) may be written as

$$P(X=x)=\frac{e^{-(x\alpha+\theta)}(1-\beta)^{\lambda x}\theta}{x!}\left\{\sum_{j=0}^{k}\binom{k}{j}(\alpha x+\theta)^{k-j}\beta^{j}(\lambda x)_{(j)}\right\},\ for\ x=1,2,... \tag{17}$$

$$P(X=0)=e^{-\theta}$$

where k=x-1 , $\alpha,\lambda,\theta>0$ and $\beta<1$.

Similarly considering Eq. (2) the pmf of LCPD-II may be written as

$$P(X=x)=A\frac{e^{-(x\alpha+\theta)}(1-\beta)^{\lambda x}}{x!}\left\{\sum_{j=0}^{x}\binom{x}{j}(\alpha x+\theta)^{x-j}\beta^{j}(\lambda x)_{(j)}\right\},for\ x=0,1,. \tag{18}$$

$= 0$, otherwise.

where $A=1-\{\alpha+\lambda\beta/(1-\beta)\}$.

(b). Cumulants of general Lagrangian distribution

According to Consul and Shenton [3], if F_r were the rth cumulants for the pgf f(z) as a function of z, and if D_r were the rth cumulants for the basic Lagrangian distribution for pgf g (z) then the cumulants of general Lagrangian distribution may be written as

$$k_1=F_1D_1,\quad k_2=F_1D_2+F_2D_1^2,\quad k_3=F_1D_3+3F_2D_1D_2+F_3D_1^3$$

$$k_4=F_1D_4+3F_3D_1^2+4F_2D_1D_3+6F_3D_1^2D_2+F_4D_1^4$$

Here D_1, D_2, D_3 and D_4 are given in Eqs. (7) - (10) respectively. Thus

$$k_1=\frac{\theta(1-\beta)}{(1-\alpha)(1-\beta)-\lambda\beta} \tag{19}$$

$$k_2=\frac{\theta(1-\beta)\{(1-\beta)^2+\lambda\beta^2}{\{(1-\alpha)(1-\beta)-\lambda\beta\}^3} \tag{20}$$

Table 2. Showing some Charlier family of Lagrange distributions of first kind.

No.	g(z)	f(z)	LCD1
1	$G_1^{\alpha,\beta,\lambda}(z)$	$G_2^{\beta,N}(z)$	$N\beta(1-\beta)^{\lambda x+N}e^{-\alpha x}(\alpha x)^{x-1}/x!$ ${}_2F_0(1-x,N+1,-\beta/\alpha x)$, $x\geq 1$
2	$G_1^{\alpha,\beta,\lambda}(z)$	$G_3^{\beta}(z)$	$A\beta(1-\beta)^{\lambda x}e^{-\alpha x}(\alpha x)^{x-1}/x!$ ${}_2F_0(1-x,\lambda x+1,-\beta/\alpha x)$, $x\geq 1$ where $A=-1/\log(1-\beta)$

where $G_1^{\alpha,\beta,\lambda}(z)$, $G_2^{\beta,N}(z)$ and $G_3^{\beta}(z)$ denote the pgf of three parameter Charlier, negative binomial and logarithmic series distribution respectively. ${}_2F_0(a,b;x)$ denotes Hypergeometric function.

Table 3. Showing some Charlier family of Lagrange distributions of second kind

No	g(z)	f(z)	LCDII
1	$G_1^{\alpha,\beta,\lambda}(z)$	$G_2^{\beta,N}(z)$	$[1-\{\alpha+\lambda\beta/(1-\beta)\}](1-\beta)^{\lambda x}e^{-\alpha x}(\alpha x)^x/x!$ $_2F_0(-x,\lambda x+N,-\beta/\alpha x)$, $x\geq 1$
2	$G_2^{\beta,N}(z)$	$G_1^{\alpha,\beta,\lambda}(z)$	$\{1-N\beta/(1-\beta)\}(1-\beta)^{Nx+\lambda}e^{-\alpha}\alpha^x/x!$ $_2F_0(-x,Nx+\lambda,-\beta/\alpha)$, $x\geq 1$

3. Conclusions

This paper defines a class of charlier type Lagrangian probability distributions by using well known Lagrange's expansions. It is also conceivable that discrete data occurring in ecology, epidemiology and meteorology can be statistically modeled on one of the distributions consider in this investigation. It may be of interest to investigate LCPD of type I and II further.

REFERENCES

1. P.C. Consul and L.R. Shenton, Use of Lagrange expression for generating generalized probability distribution. SIAM Journal of Applied Mathematics. 23 239-248 (1972).
2. P.C. Consul and L.R. Shenton, Some interesting properties of Lagrangian distributions. Communications in Statistics. 2 263-272 (1973).
3. P.C. Consul and L.R. Shenton, On the probabilistic structure and properties of discrete Lagrangian distributions. Statistical Distribution in Scientific Work 1: Models and Structure, (Edited by G.P.Patil, S.Kotz and J.K.Kotz), 41-48, Dordrecht Reidel, (1975)
4. S.G. Mohanty, On generalized two-coin tossing problem, Biometrische Zeitschrift, 8 266-272 (1966).
5. P.C. Consul and G.C. Jain, A generalization of Poisson distribution, Technometrics, 15 791-799 (1973).
6. P.C. Consul, Lagrange and related probability distribution, Encyclopedia of Statistical Sciences, (Edited by S. Kotz, N.L. Johnson and C.B. Read), 4 448-454, New York: Wiley, (1983)
7. K.G. Janardan and B.R. Rao, Lagrange distribution of second kind and weighted distributions, SIAM Journal of Applied Mathematics, 43 302-313, (1983).
8. G.C. Jain and R.P.Gupta, A Logarithmic series type distribution, Trabajos de Estadistica, 24, 99-105, (1973).
9. J. Medhi and M. Borah, On generalized Gegenbauer polynomials and associated probabilities, Sankhya, ser. B, 46,157-165, (1984).
10. M. Borah and R. Beguem, A class of Charlier family of Lagrangian probability distributions, proceeding of Math. And Stat. In Engineering and Technology, A. Chattopadhyay (editor), (1997).
11. C.B. Williams, The number of publications written by biologists, Annals of Eugenics, 12, 143-146, (1944).
12. G.C. Jain, A power series distribution associated with Lagrange expression, Biometrische Zeitschrift, 17, 85-97, (1975).
